Statistical Inference I

GLM, Contrasts & RFT

Guillaume Flandin

Wellcome Trust Centre for Neuroimaging

University College London

SPM Course

Lausanne, November 2013

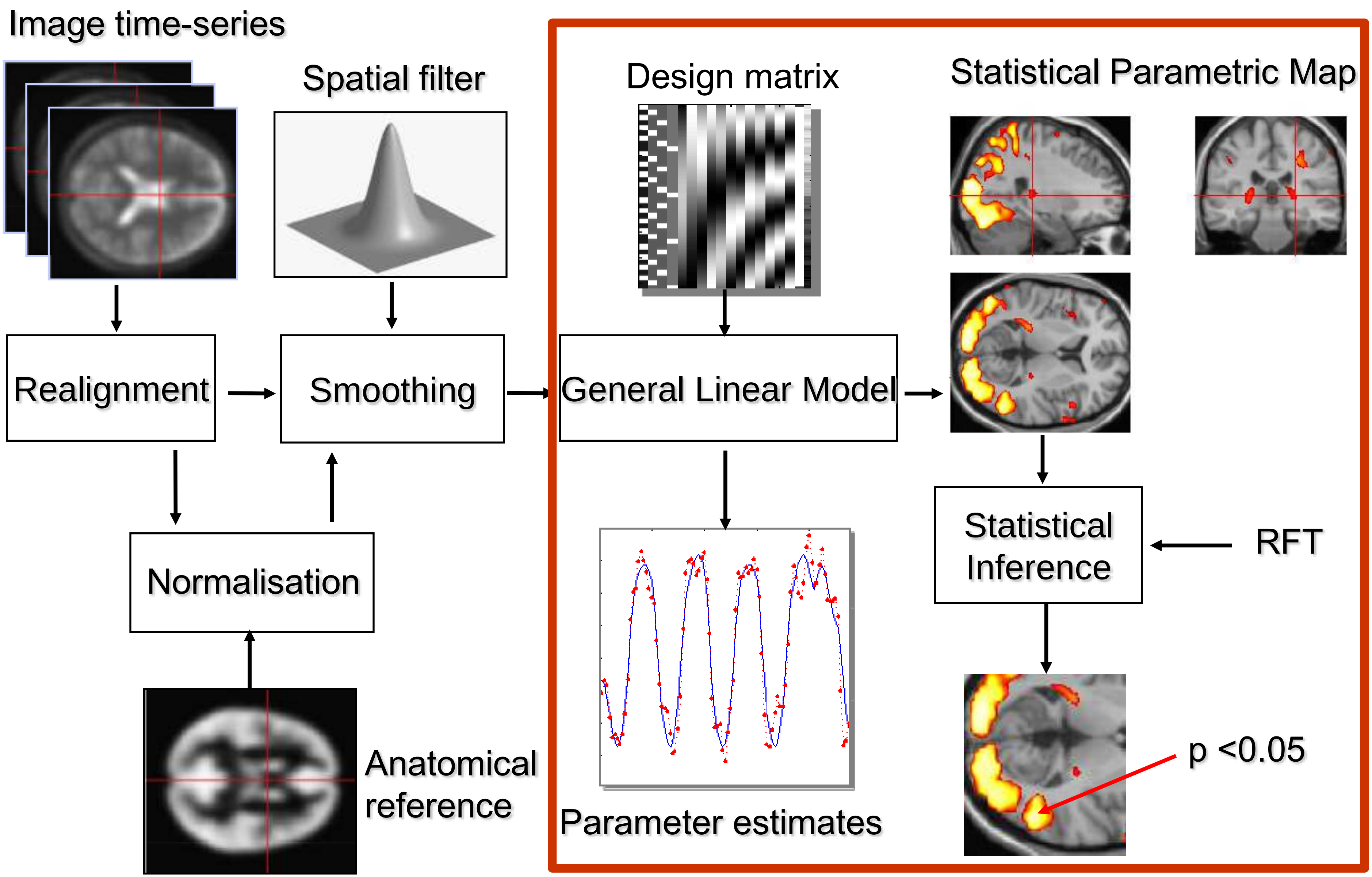
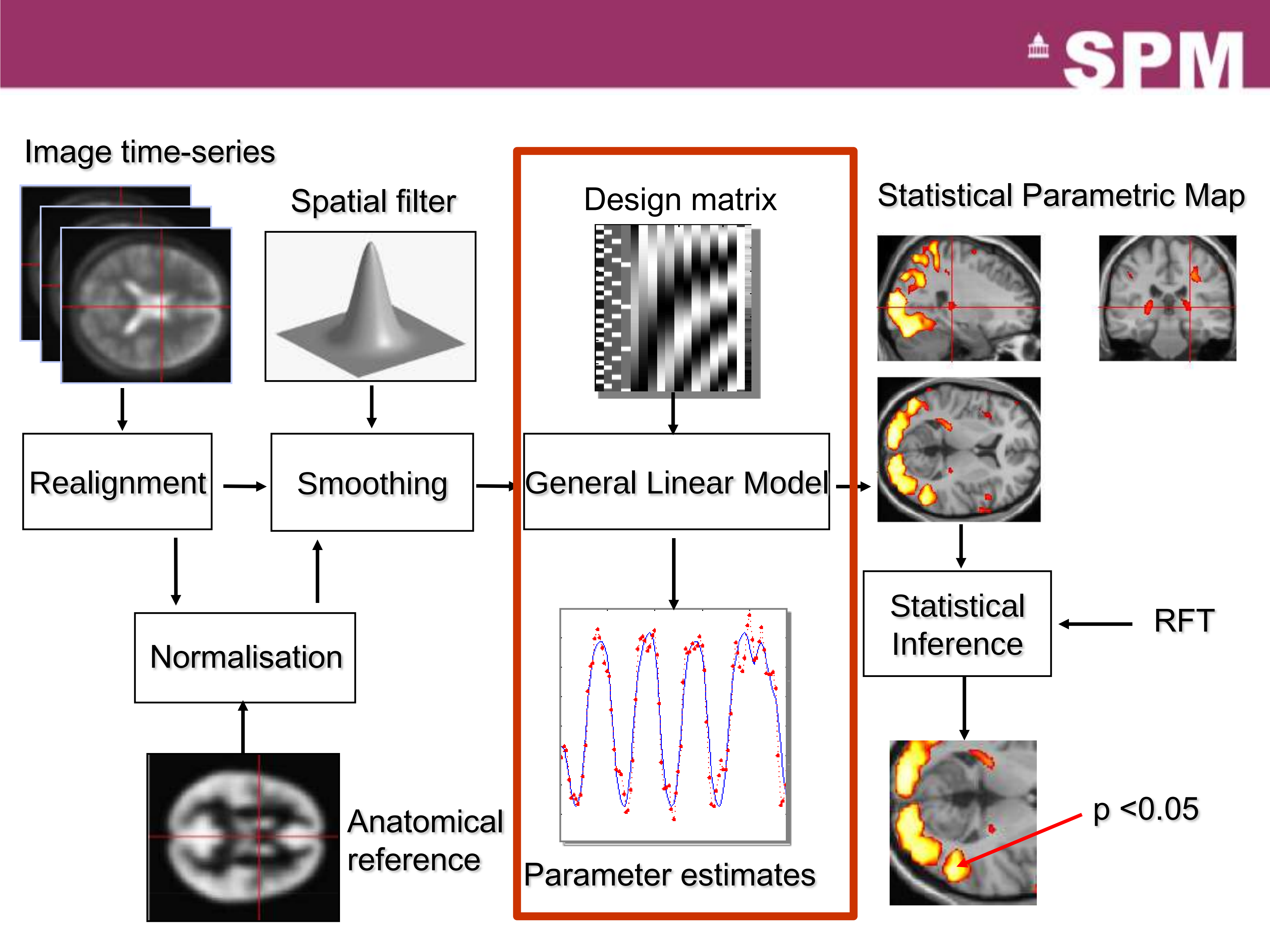


Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Statistical
Inference
RFT
Normalisation
Anatomical
reference
Parameter estimates
p <0.05



SPM
Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Normalisation
Anatomical reference
Parameter estimates
Statistical Inference
RFT
p <0.05

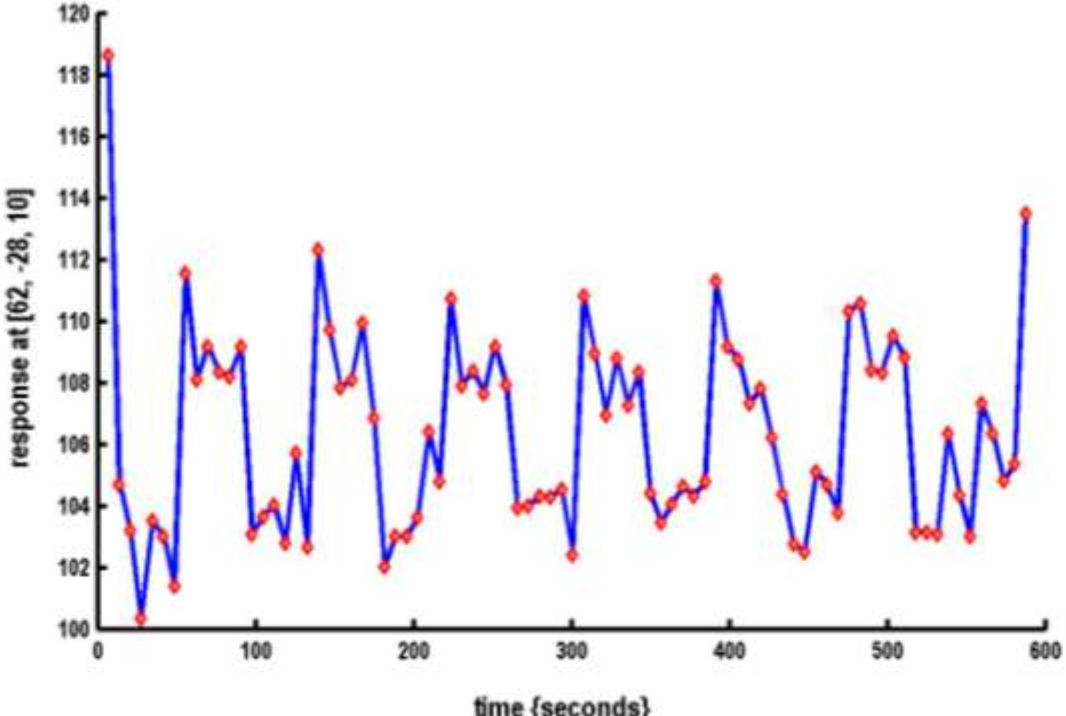
fMRI experiment example

One session

Passive word listening versus rest

7 cycles of rest and listening

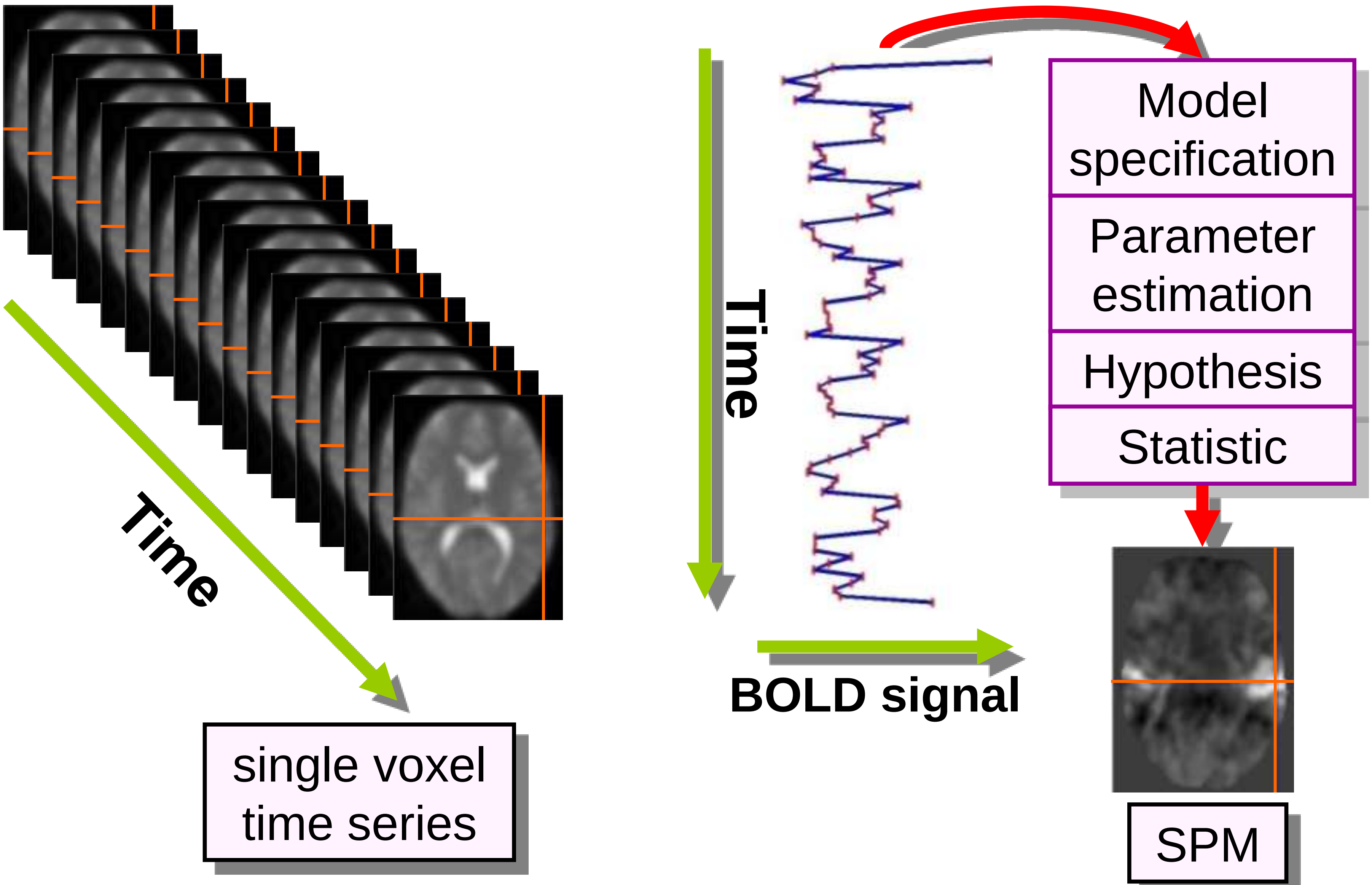
Blocks of 6 scans with 7 sec TR



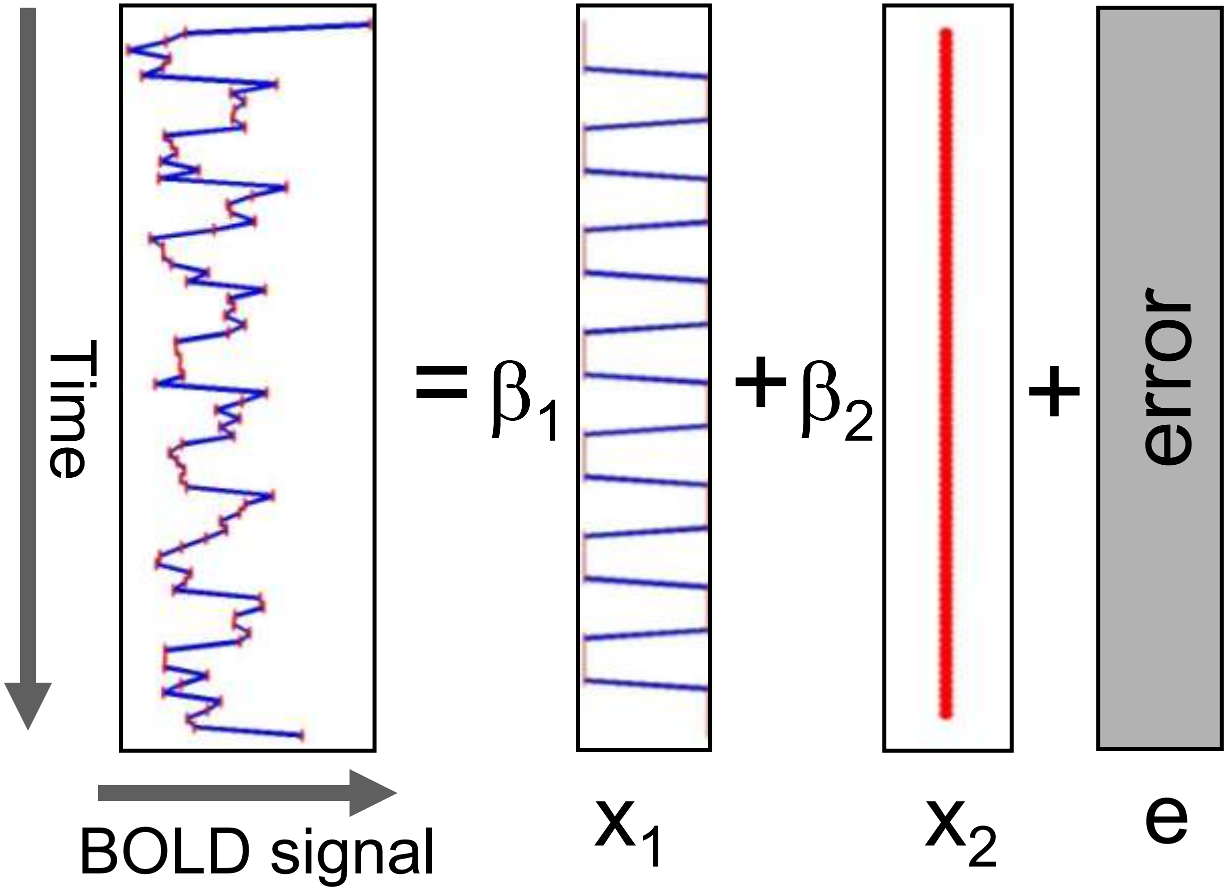
Stimulus function

Question: Is there a change in the BOLD response between listening and rest?

Voxel-wise time series analysis

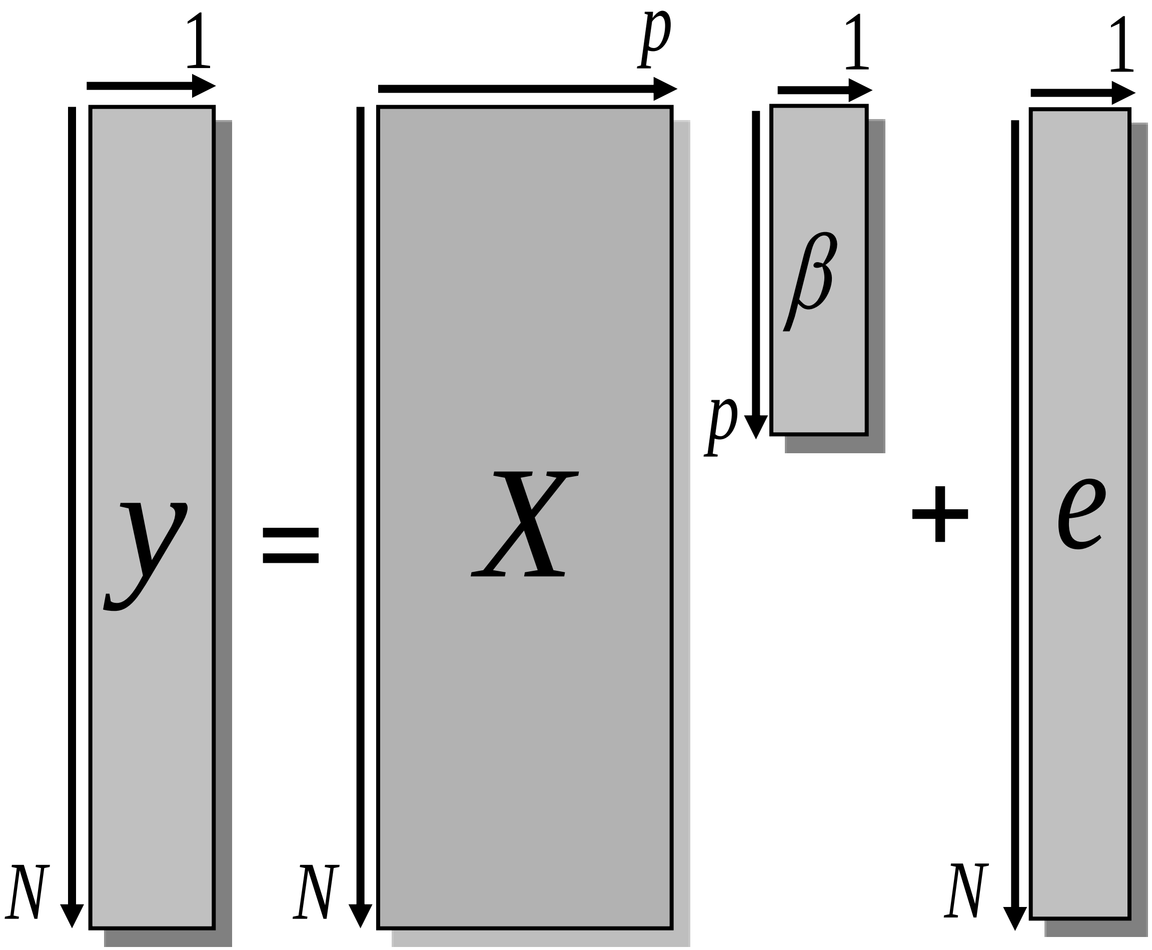


Single voxel regression model



$$y = \beta_1 x_1 + \beta_2 x_2 + e$$

Mass-univariate analysis: voxel-wise GLM



$$y = X\beta + e$$

$$e \sim N(0, \sigma^2 I)$$

Model is specified by
1. Design matrix X
2. Assumptions about e

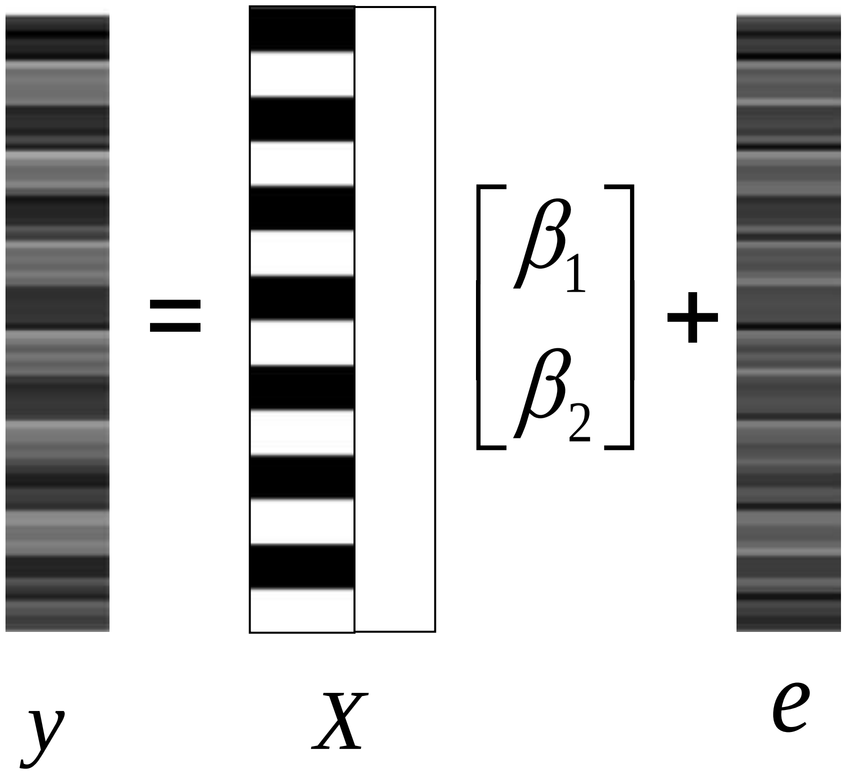
N: number of scans
p: number of regressors

The design matrix embodies all available knowledge about experimentally controlled factors and potential confounds.

GLM: a flexible framework for parametric analyses

- one sample *t*-test
- two sample *t*-test
- paired *t*-test
- Analysis of Variance (ANOVA)
- Analysis of Covariance (ANCoVA)
- correlation
- linear regression
- multiple regression

Parameter estimation



$$y = X\beta + e$$

Objective: estimate parameters to minimize $\sum_{t=1}^{N} e_t^2$

Ordinary least squares estimation (OLS) (assuming i.i.d. error):

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Problems of this model with fMRI time series

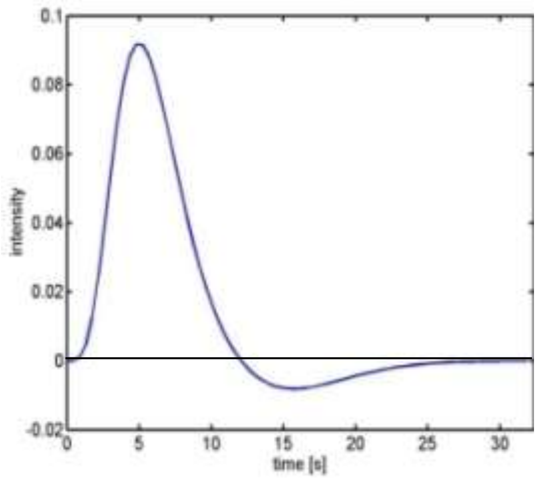
1. The *BOLD response* has a delayed and dispersed shape.

2. The BOLD signal includes substantial amounts of *low-frequency noise* (eg due to scanner drift).

3. Due to breathing, heartbeat & unmodeled neuronal activity, the *errors are serially correlated*. This violates the assumptions of the noise model in the GLM.

Problem 1: BOLD response

Hemodynamic response function (HRF):



Linear time-invariant (LTI) system:

$$u(t) \longrightarrow \boxed{hrf(t)} \longrightarrow x(t)$$

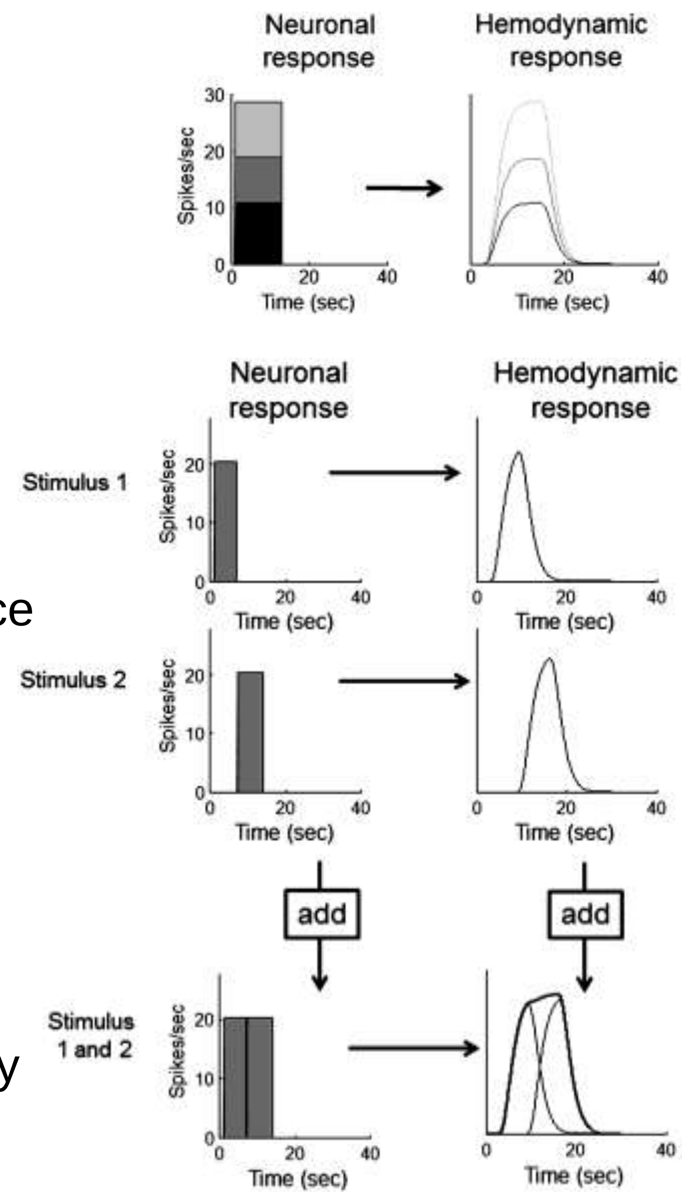
Convolution operator:

$$\begin{aligned} x(t) &= u(t) * hrf(t) \\ &= \int_0^t u(\tau)\,hrf(t-\tau)\,d\tau \end{aligned}$$

Scaling

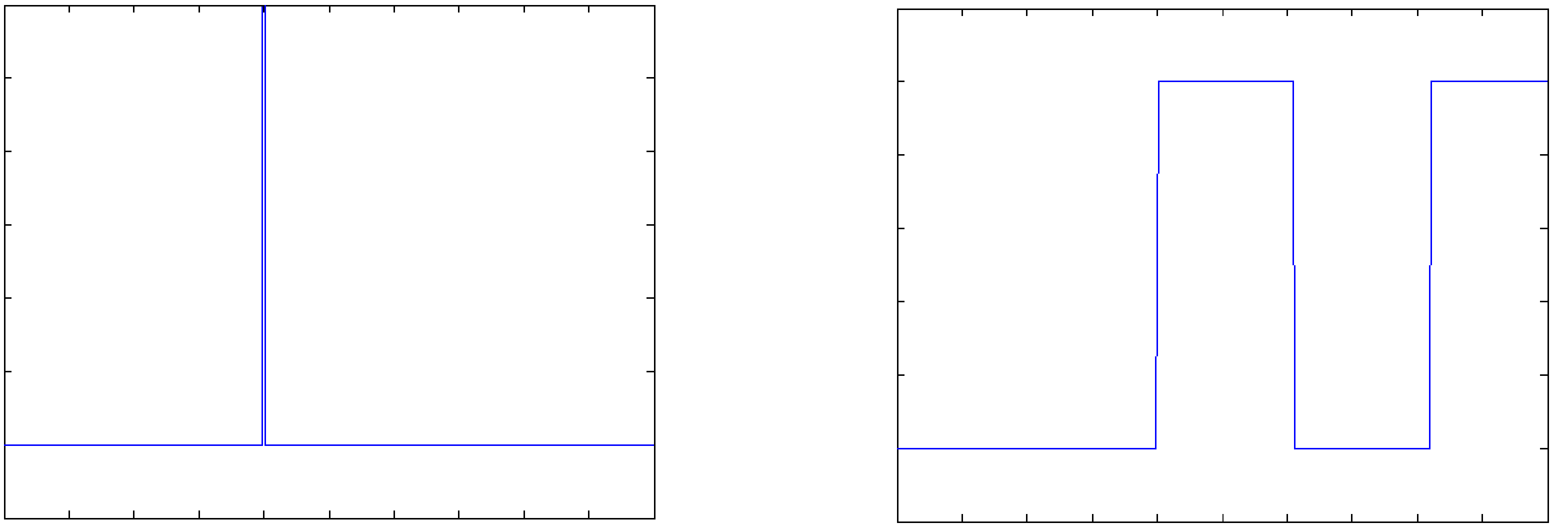
Shift invariance

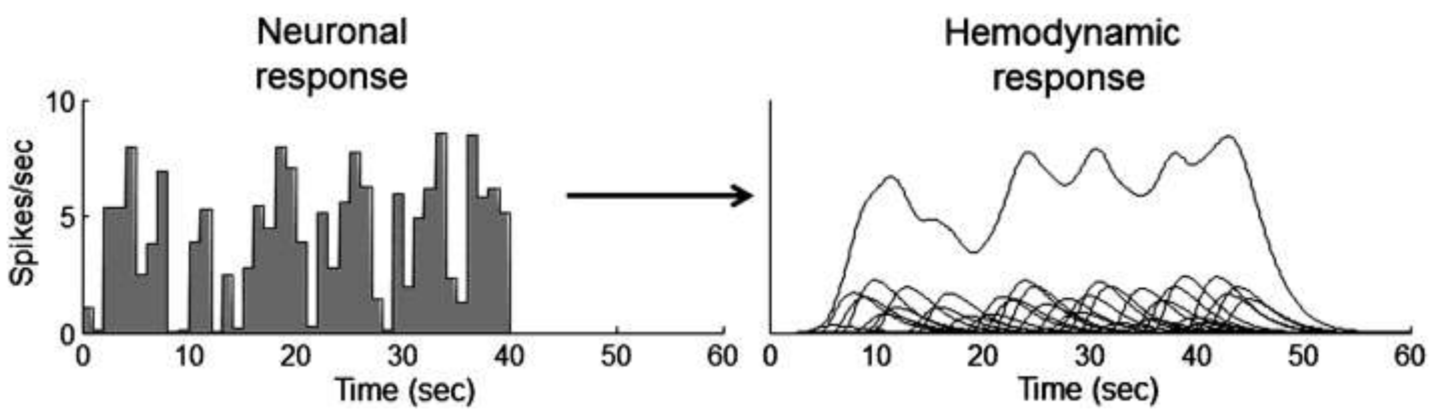
Additivity



Boynton et al, NeuroImage, 2012.

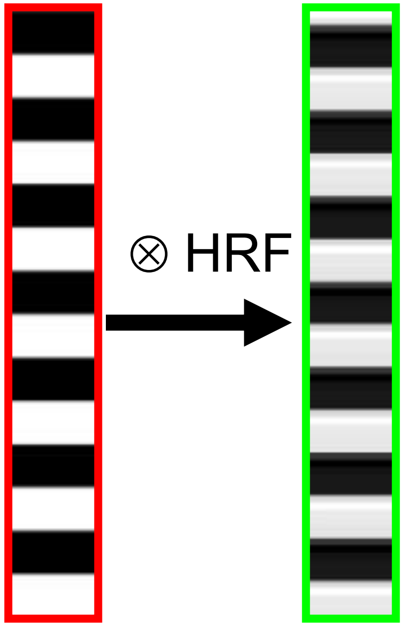
Problem 1: BOLD response
Solution: Convolution model



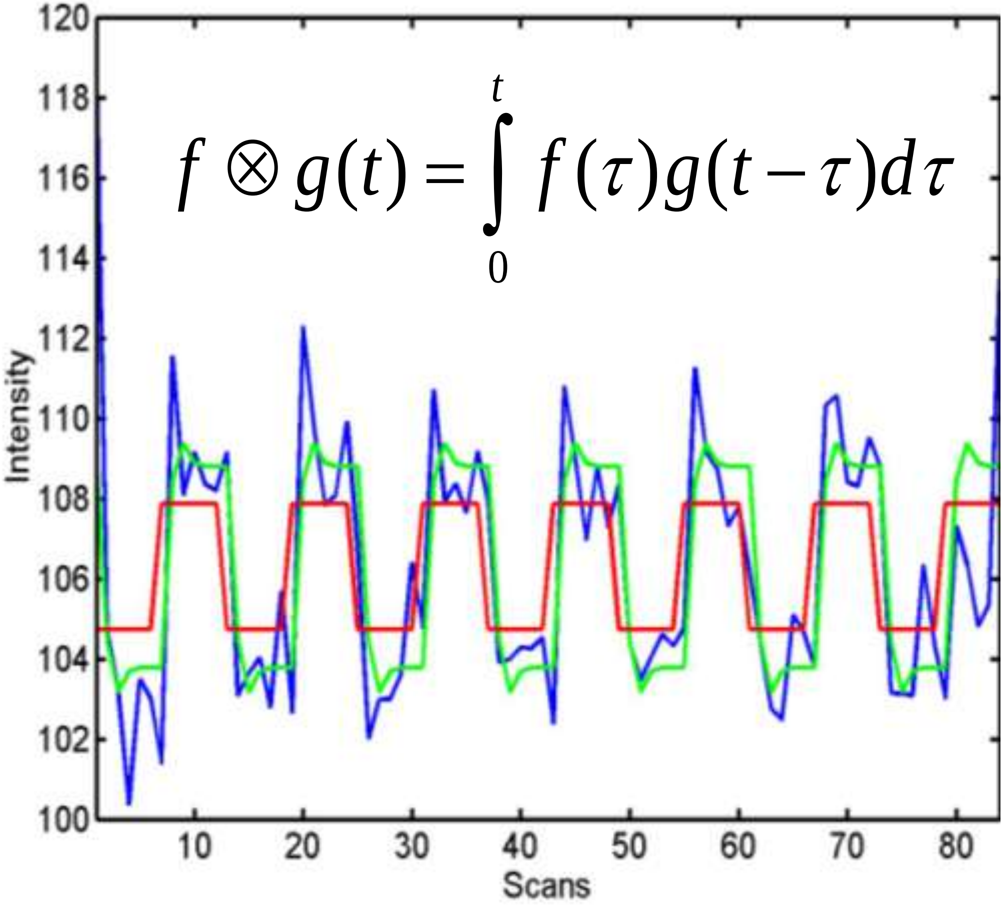


Convolution model of the BOLD response

Convolve stimulus function with a canonical hemodynamic response function (HRF):

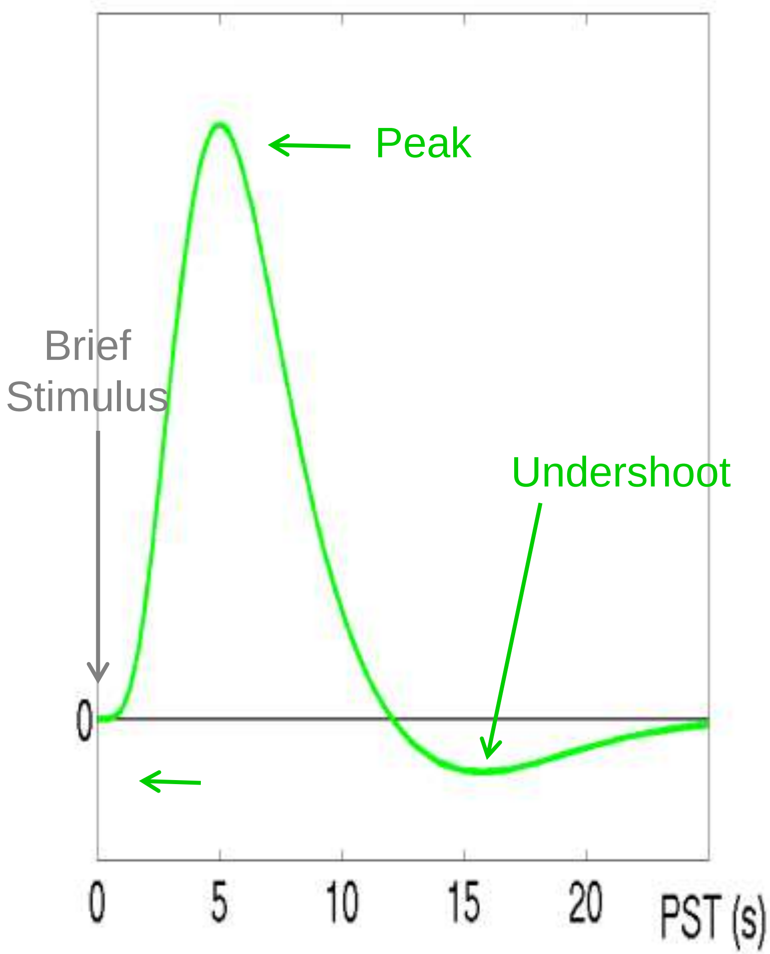


$$f \otimes g(t) = \int_0^t f(\tau)g(t-\tau)d\tau$$

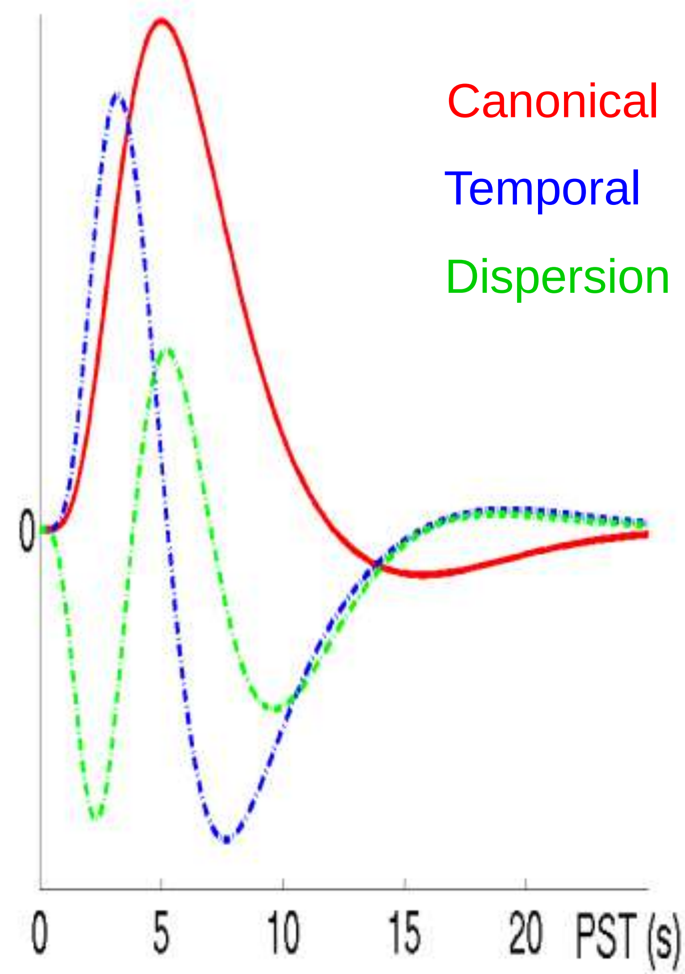


Hemodynamic Response ⇨ Temporal Basis Set

Canonical HRF

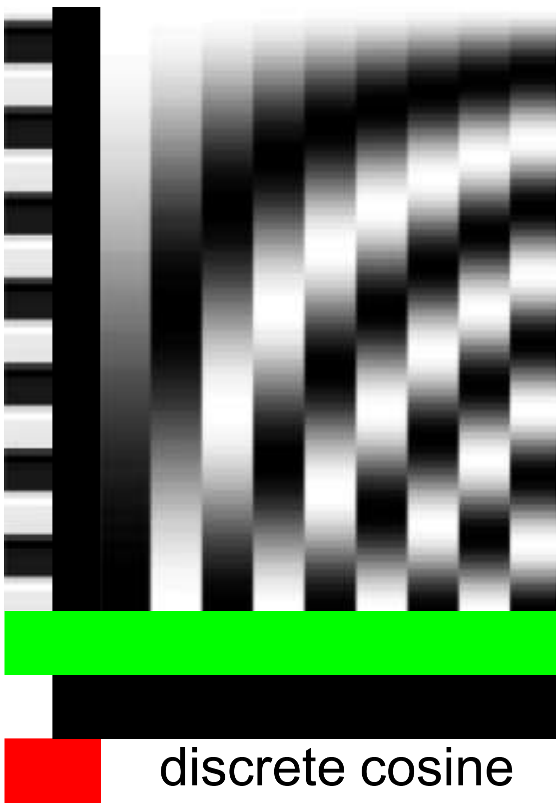


Informed Basis Set

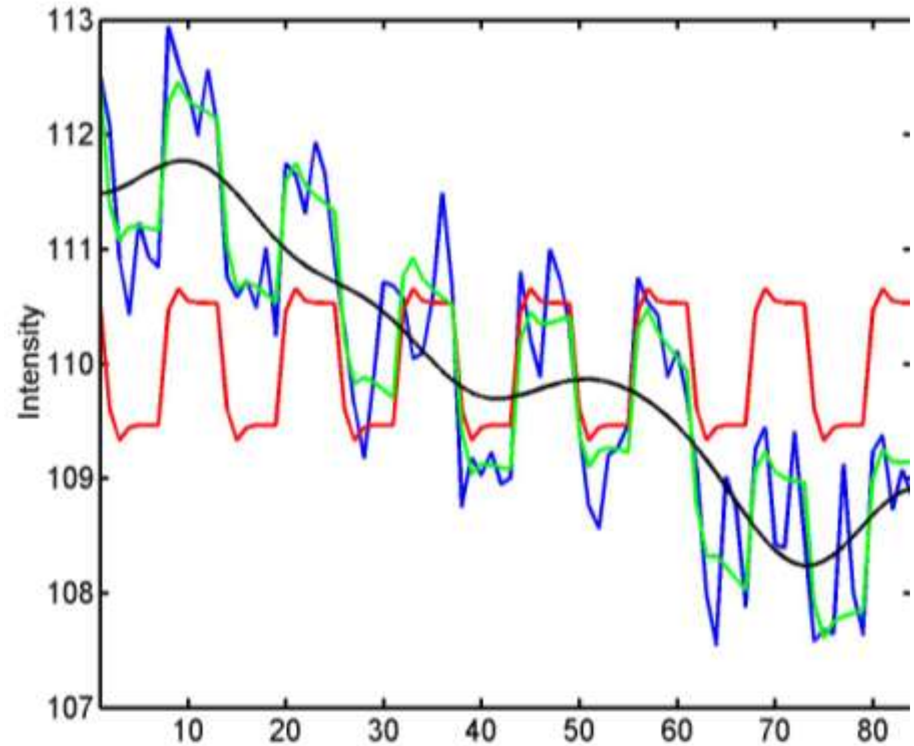


Problem 2: Low-frequency noise

Solution: High pass filtering



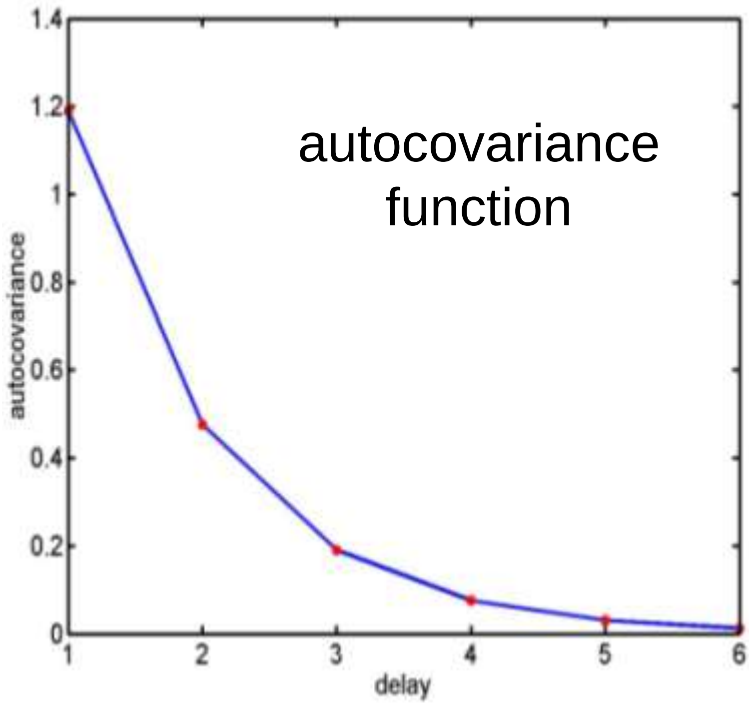
discrete cosine transform (DCT) set

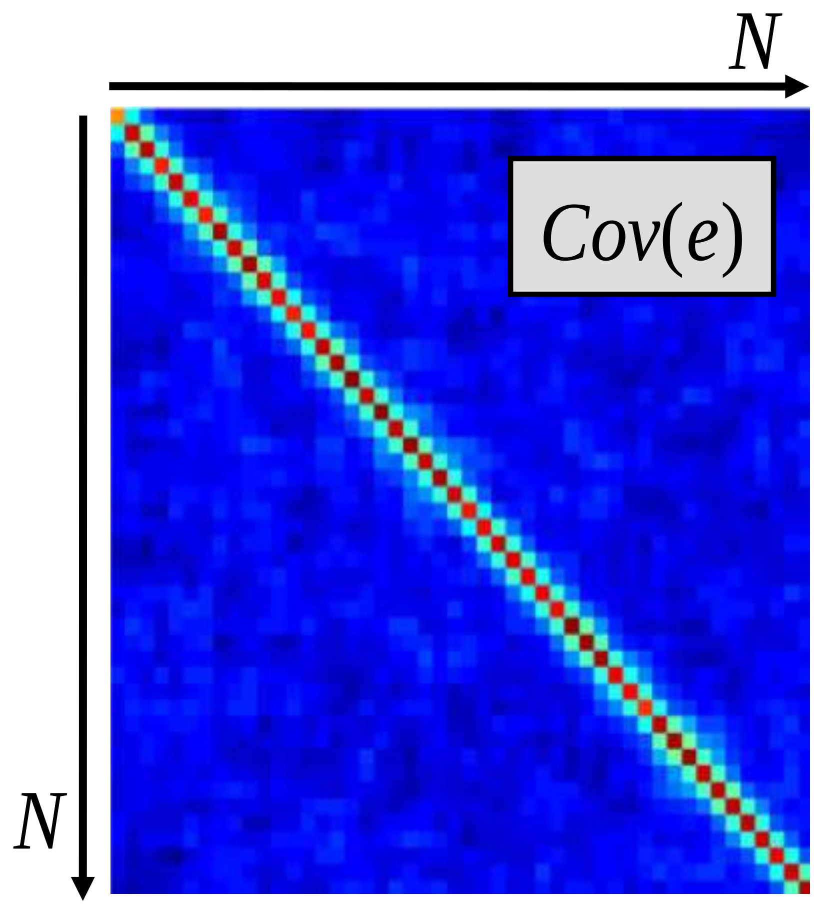


blue =	data
black =	mean + low-frequency drift
green = account	predicted response, taking into low-frequency drift
red = into	predicted response, NOT taking account low-frequency drift

Problem 3: Serial correlations

i.i.d: $\cancel{e \sim N(0, \sigma^2 I)}$





Multiple covariance components

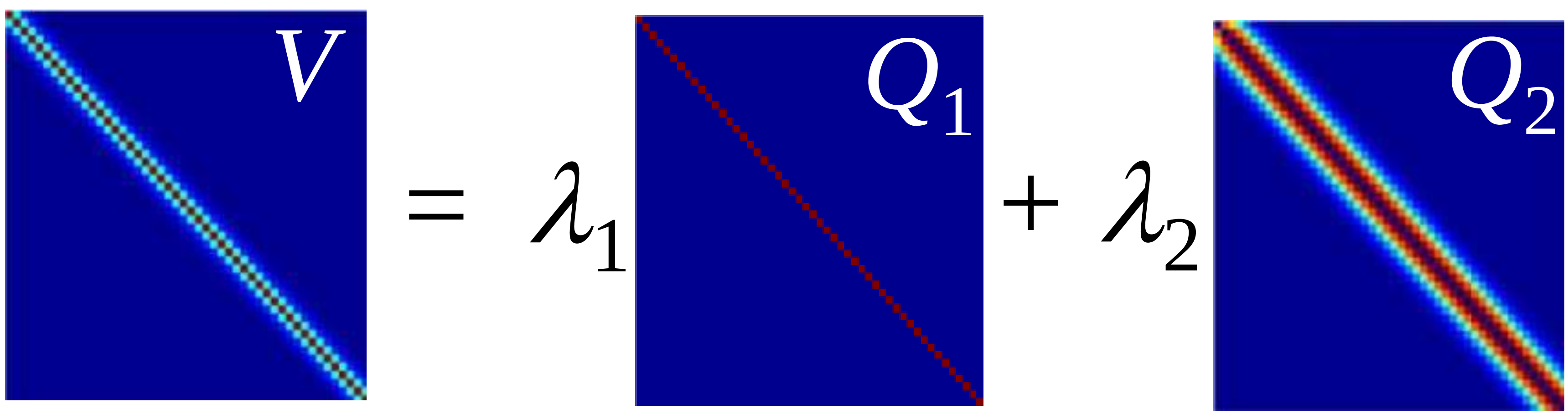
enhanced noise model at voxel *i*

$$e_i \sim N(0, C_i)$$

$$C_i = \sigma_i^2 V$$

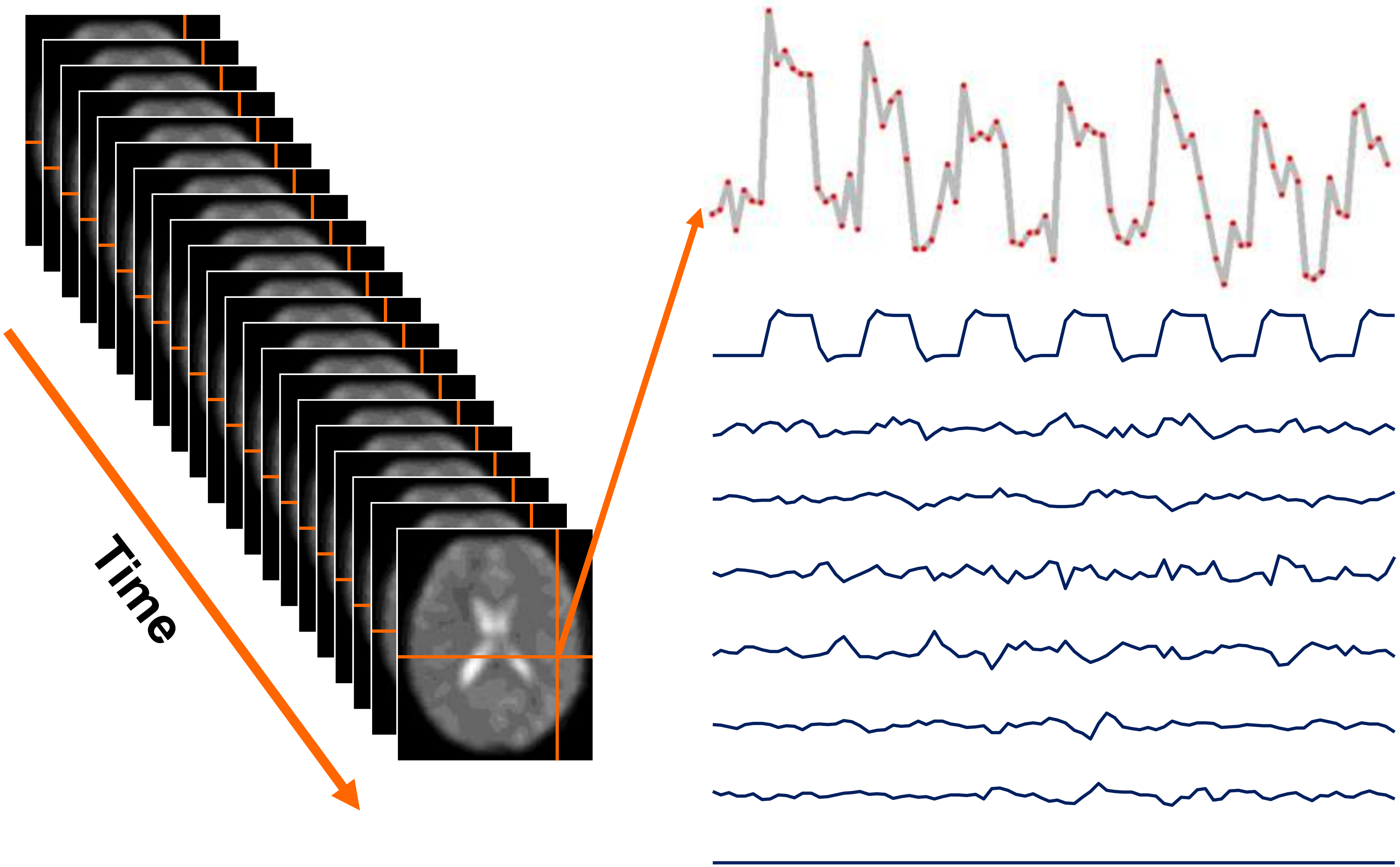
$$V = \sum \lambda_j Q_j$$

error covariance components Q
and hyperparameters λ

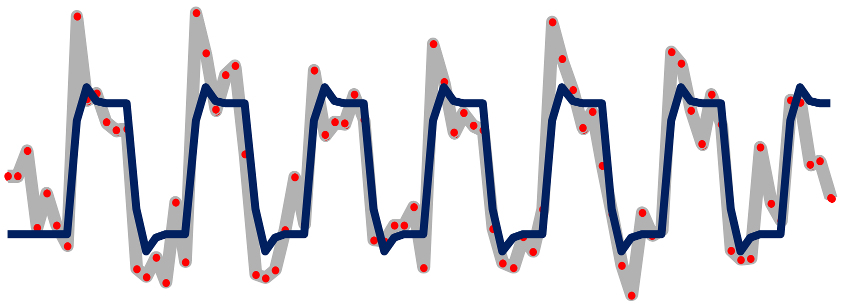


Estimation of hyperparameters λ with ReML (Restricted Maximum Likelihood).

Summary: a mass-univariate approach



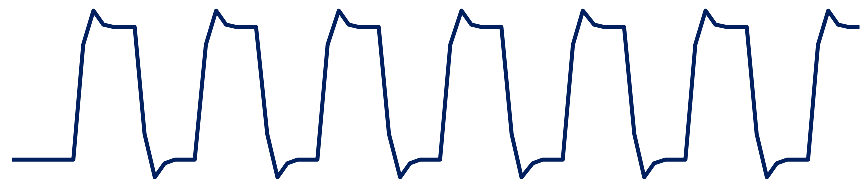
Summary: Estimation of the parameters



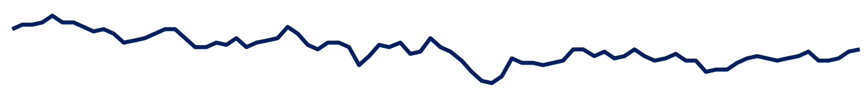
i.i.d. assumptions: $\varepsilon \sim N(0, \sigma^2 I)$

OLS estimates: $\hat{\beta} = (X^T X)^{-1} X^T y$

$\hat{\beta}_1 = 3.9831$

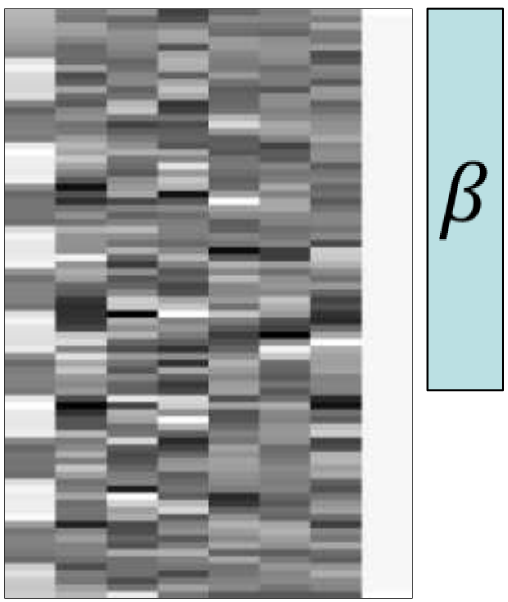


$\hat{\beta}_{2-7} = \{0.6871, 1.9598, 1.3902, 166.1007, 76.4770, -64.8189\}$



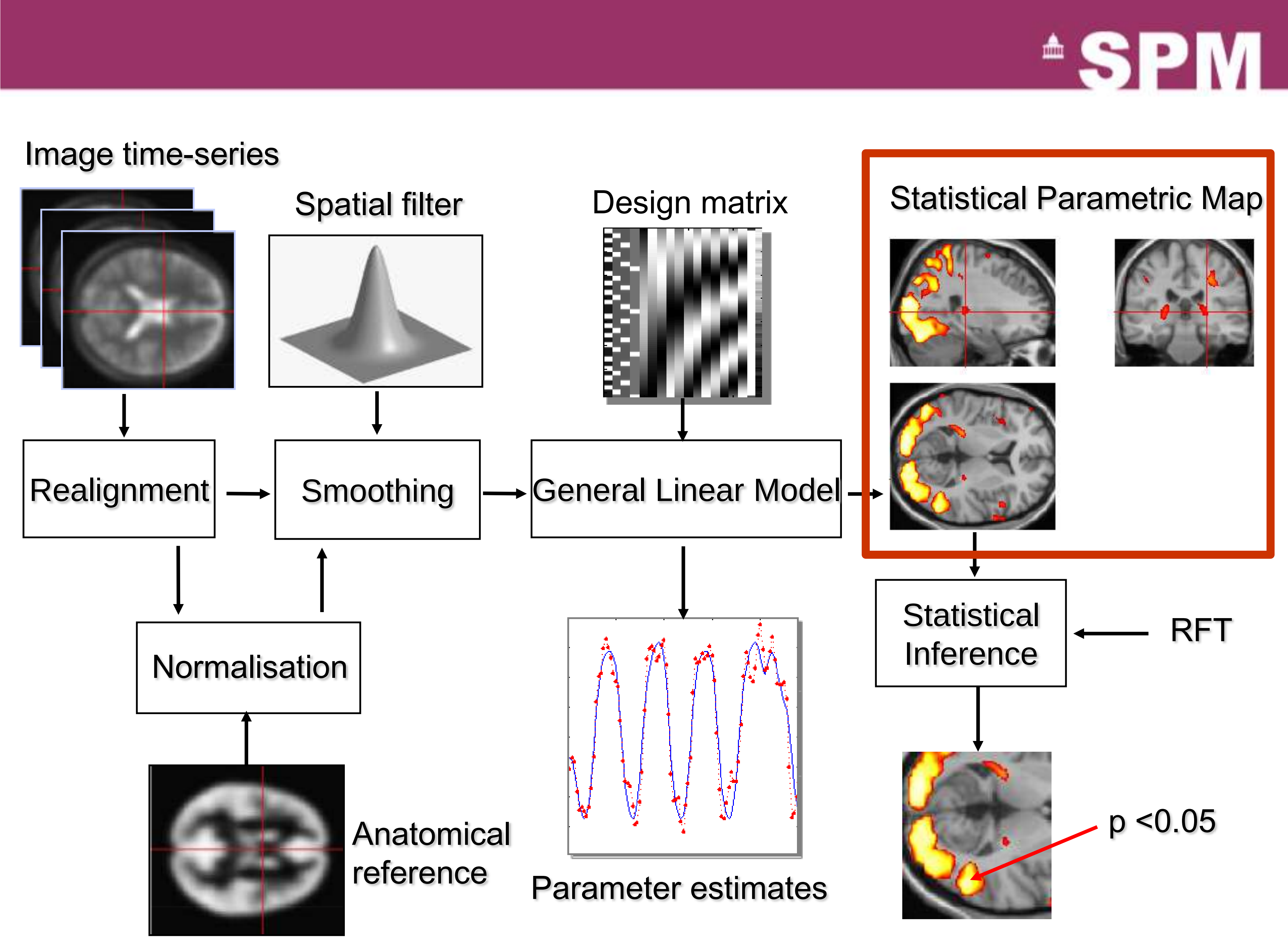
$\hat{\beta}_8 = 131.0040$

$y =$



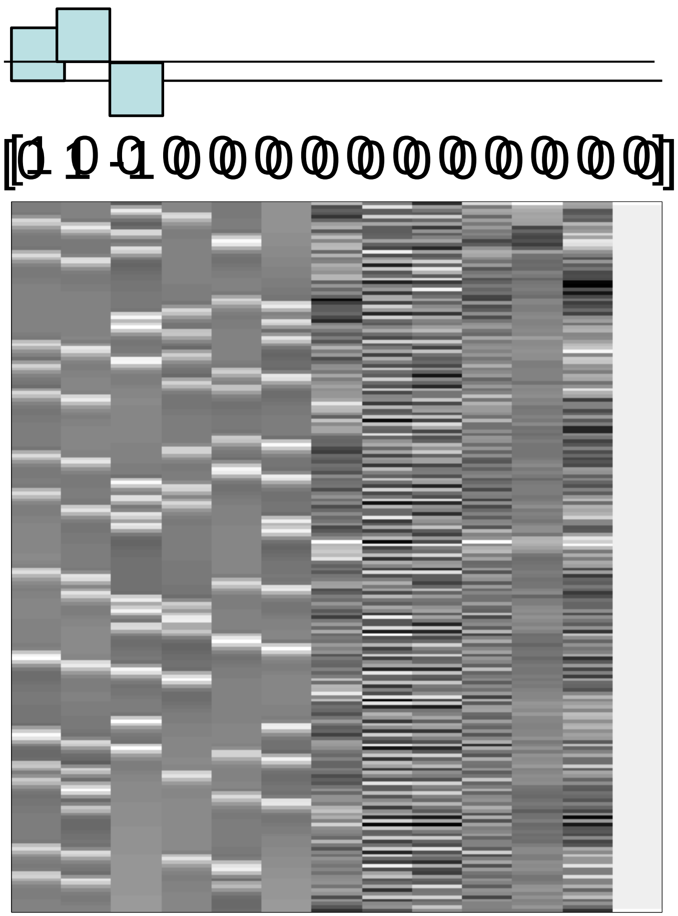
$+\varepsilon$

$\hat{\varepsilon} =$



SPM
Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Statistical Inference
RFT
Normalisation
Anatomical reference
Parameter estimates
p <0.05

Contrasts



- A contrast selects a specific effect of interest.
  - ⇨ A contrast c is a vector of length p.
  - ⇨ $c^T\beta$ is a linear combination of regression coefficients β.

$$c = [1\ 0\ 0\ 0\ \ \ldots]^T$$

$$c^T\beta = \mathbf{1} \times \beta_1 + \mathbf{0} \times \beta_2 + \mathbf{0} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$
$$= \boldsymbol{\beta_1}$$

$$c = [0\ 1\ -1\ 0\ \ \ldots]^T$$

$$c^T\beta = \mathbf{0} \times \beta_1 + \mathbf{1} \times \beta_2 + \mathbf{-1} \times \beta_3 + \mathbf{0} \times \beta_4 + \cdots$$
$$= \boldsymbol{\beta_2 - \beta_3}$$

Hypothesis Testing

To test an hypothesis, we construct "test statistics".

- **Null Hypothesis H_0**

  Typically what we want to disprove (no effect).

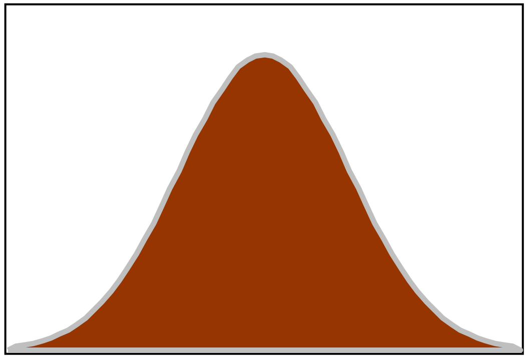
  ⇨ The Alternative Hypothesis H_A expresses outcome of interest.

- **Test Statistic T**

  The test statistic summarises evidence about H_0.

  Typically, test statistic is small in magnitude when the hypothesis H_0 is true and large when false.

  ⇨ We need to know the distribution of T under the null hypothesis.

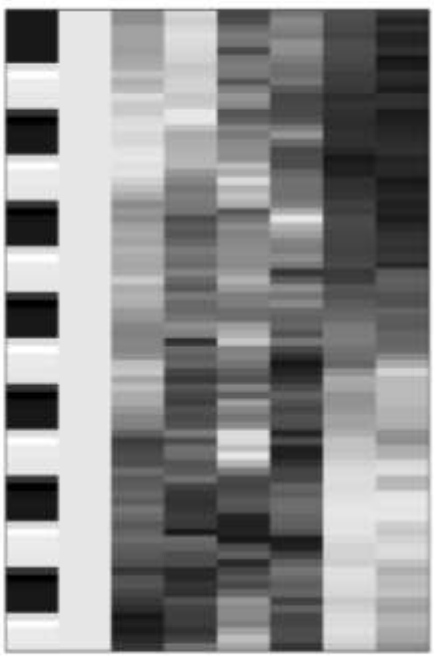


Null Distribution of T

***T*-test** - one dimensional contrasts – SPM{*t*}

c^T = 1 0 0 0 0 0 0 0

$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5$ ...



Question: box-car amplitude > 0 ?

=

$\beta_1 = c^T\beta > 0$?

Null hypothesis: $H_0: c^T\beta = 0$

Test statistic:

$$T = \frac{\textbf{\textit{contrast} of estimated parameters}}{\sqrt{\textbf{variance estimate}}}$$

$$T = \frac{c^T\hat{\beta}}{\sqrt{\mathrm{var}(c^T\hat{\beta})}} = \frac{c^T\hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T\left(X^TX\right)^{-1}c}} \sim t_{N-p}$$

***T*-test:** one dimensional contrasts – SPM{*t*}

- *T*-test is a *signal-to-noise* measure (ratio of estimate to standard deviation of estimate).

- Alternative hypothesis:

$$H_0: \; c^T\beta = 0 \qquad \text{vs} \qquad H_A: \; c^T\beta > 0$$

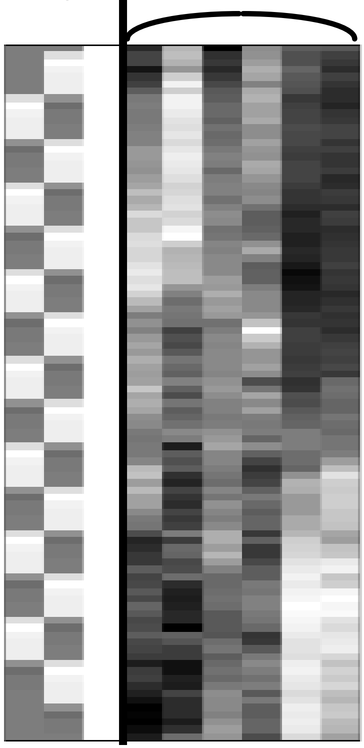
- *T*-contrasts are simple combinations of the betas; the T-statistic does not depend on the scaling of the regressors or the scaling of the contrast.

***F*-test** - the extra-sum-of-squares principle

- Model comparison:

Null Hypothesis $\mathbf{H_0}$: True model is X_0 (reduced model)

$\boldsymbol{X_0}$ $\boldsymbol{X_1}$

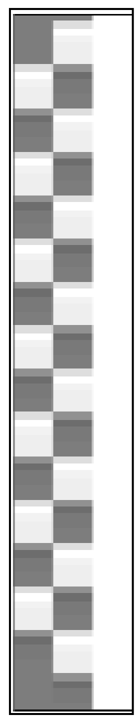


Full model ?

RSS

$$\sum \hat{\varepsilon}^2_{full}$$

$\boldsymbol{X_0}$



or Reduced model?

$\mathbf{RSS_0}$

$$\sum \hat{\varepsilon}^2_{reduced}$$

Test statistic: ratio of explained variability and unexplained variability (error)

$$F \propto \frac{RSS_0 - RSS}{RSS}$$

$$F \propto \frac{ESS}{RSS} \sim F_{\nu_1,\nu_2}$$

ν_1 = rank(X) – rank(X_0)
ν_2 = N – rank(X)

***F*-test** - multidimensional contrasts – SPM{*F*}

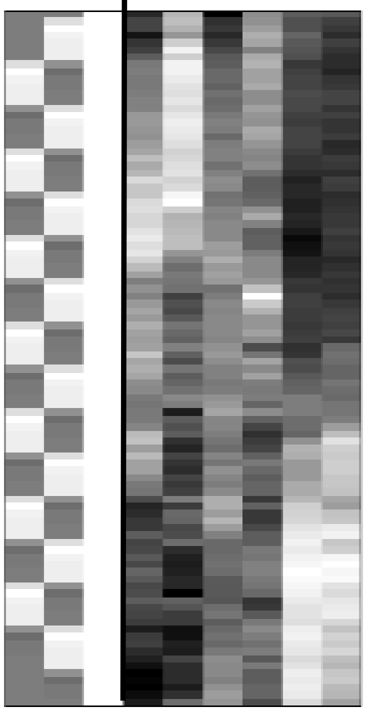
- Tests multiple linear hypotheses:

H_0: True model is X_0

H_0: $\beta_4 = \beta_5 = \dots = \beta_9 = 0$

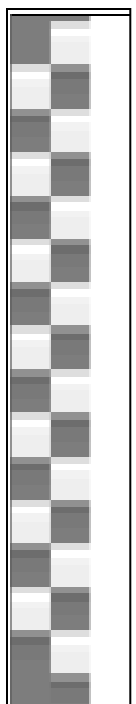
test H_0 : $c^T\beta = 0$?

X_0 | X_1 $(\beta_{4\text{-}9})$



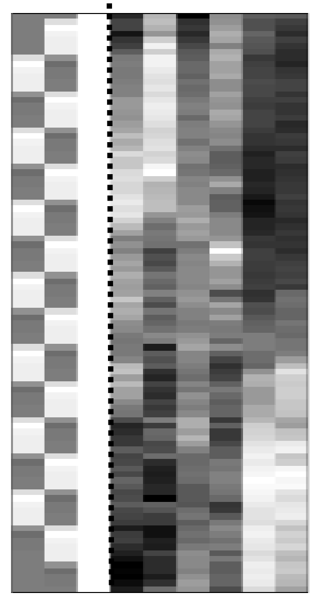
Full model?

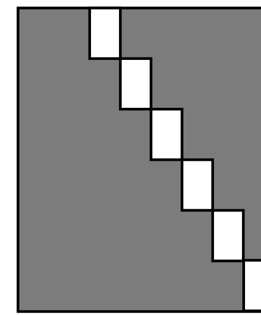
X_0

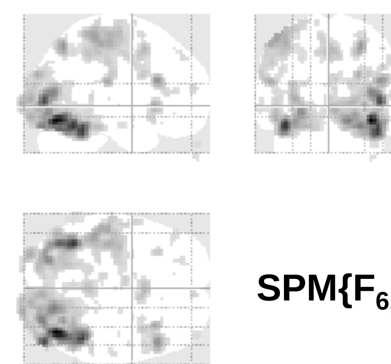


Reduced model?

$$c^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

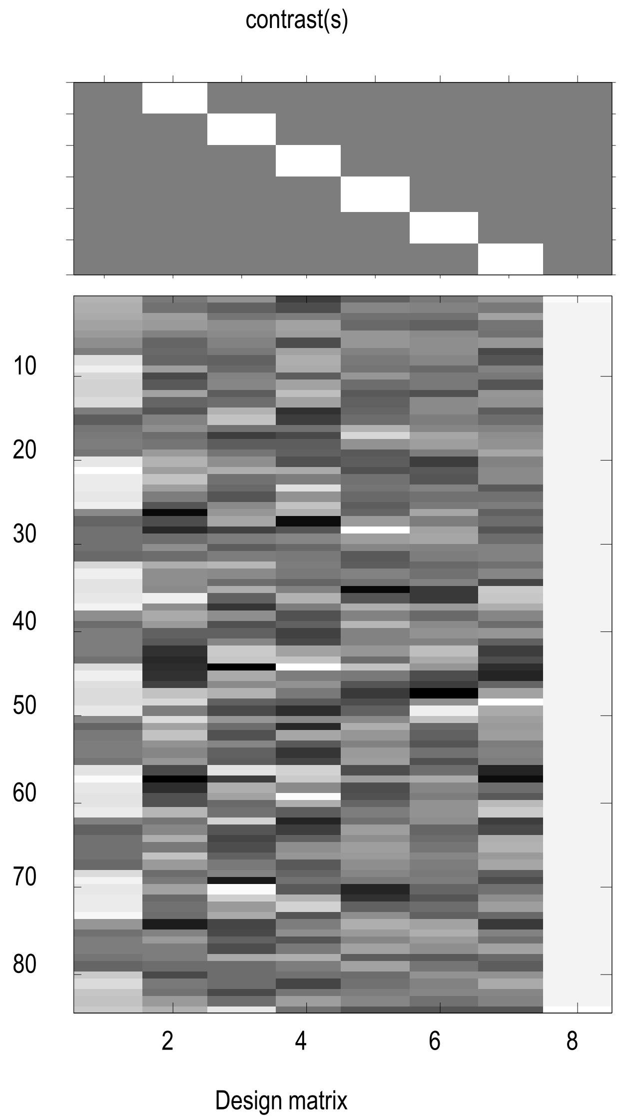


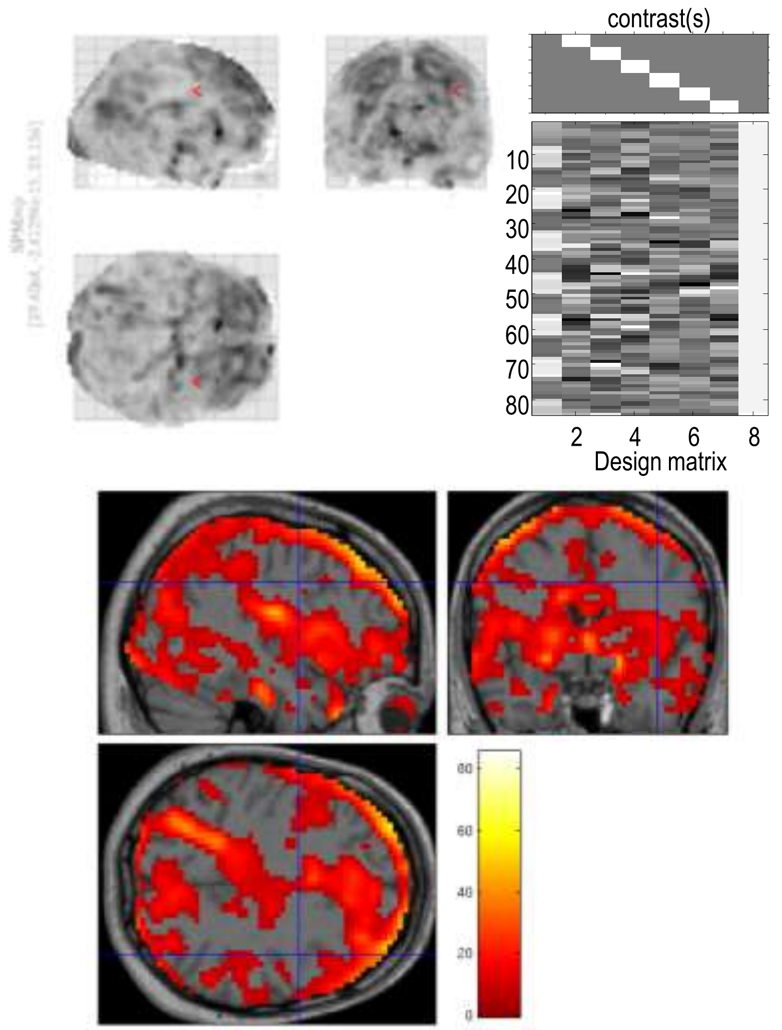




SPM{$F_{6,322}$}

***F*-test** example: movement related effects





F-test: summary

- F-tests can be viewed as testing for the additional variance explained by a larger model wrt a simpler (***nested***) model ⇨ ***model comparison***.

- F tests a weighted **sum of squares** of one or several combinations of the regression coefficients β.

- Hypotheses:

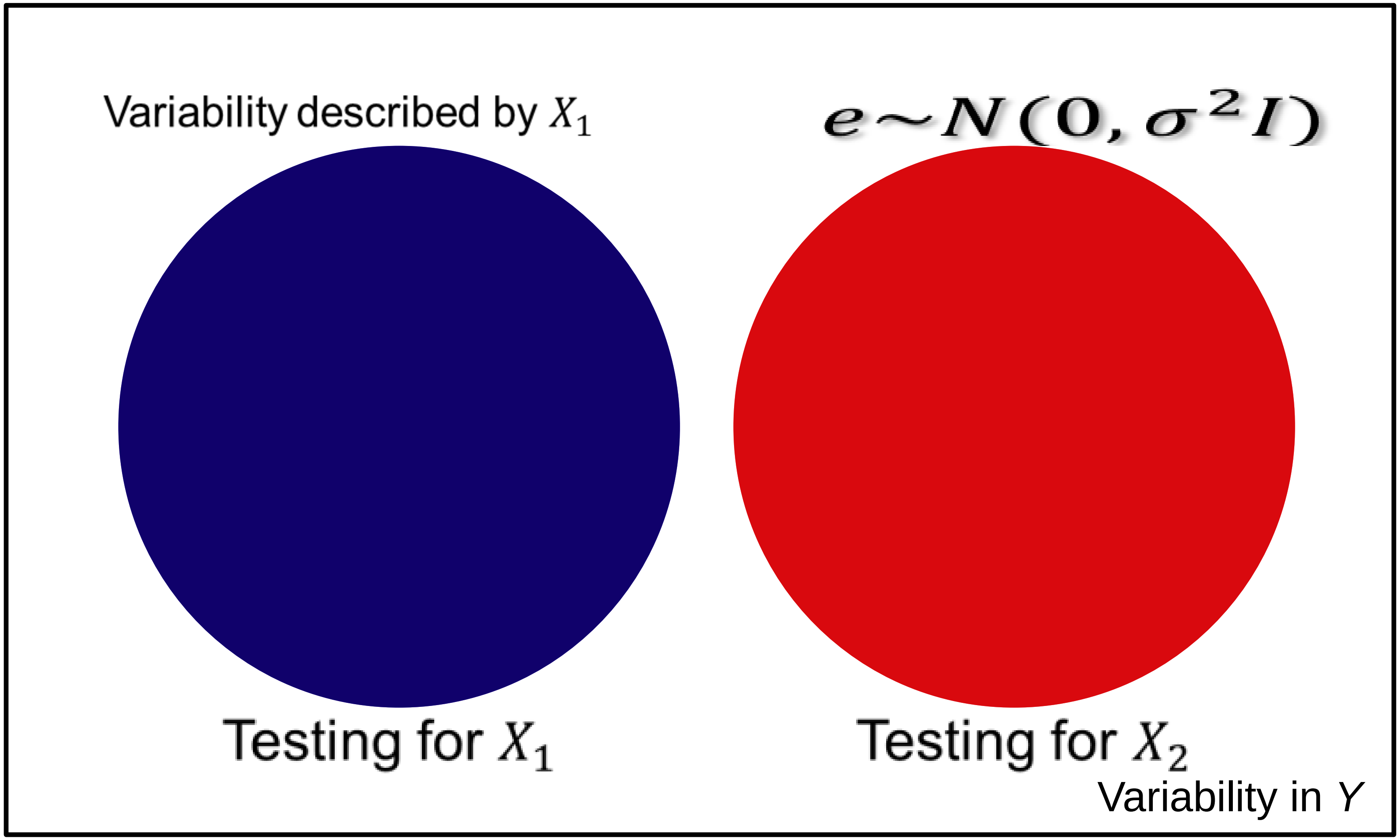
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{Null Hypothesis } H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

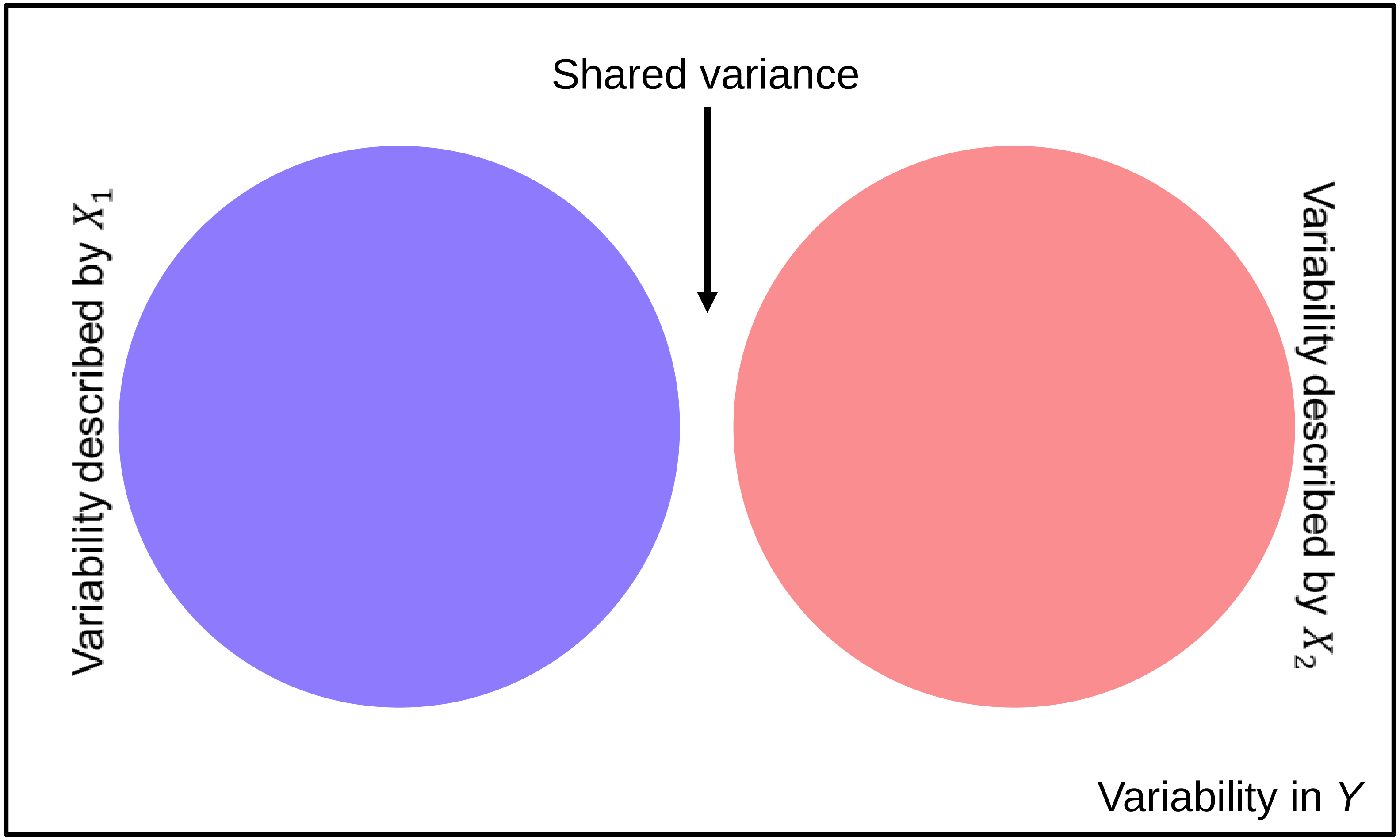
$$\text{Alternative Hypothesis } H_A : \text{at least one } \beta_k \neq 0$$

- In testing uni-dimensional contrast with an *F*-test, for example $\beta_1 - \beta_2$, the result will be the same as testing $\beta_2 - \beta_1$. It will be exactly the square of the *t*-test, testing for both positive and negative effects.

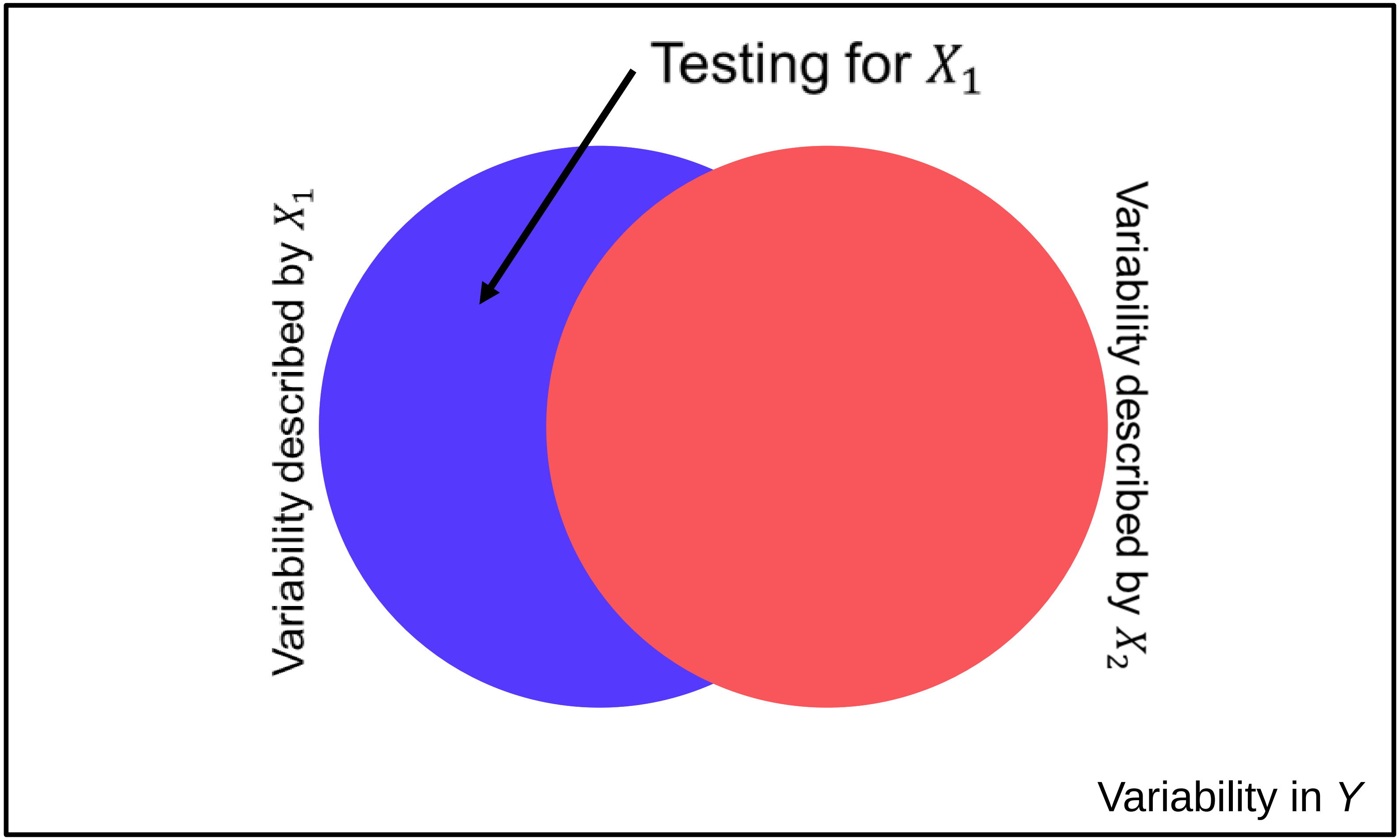
Orthogonal regressors



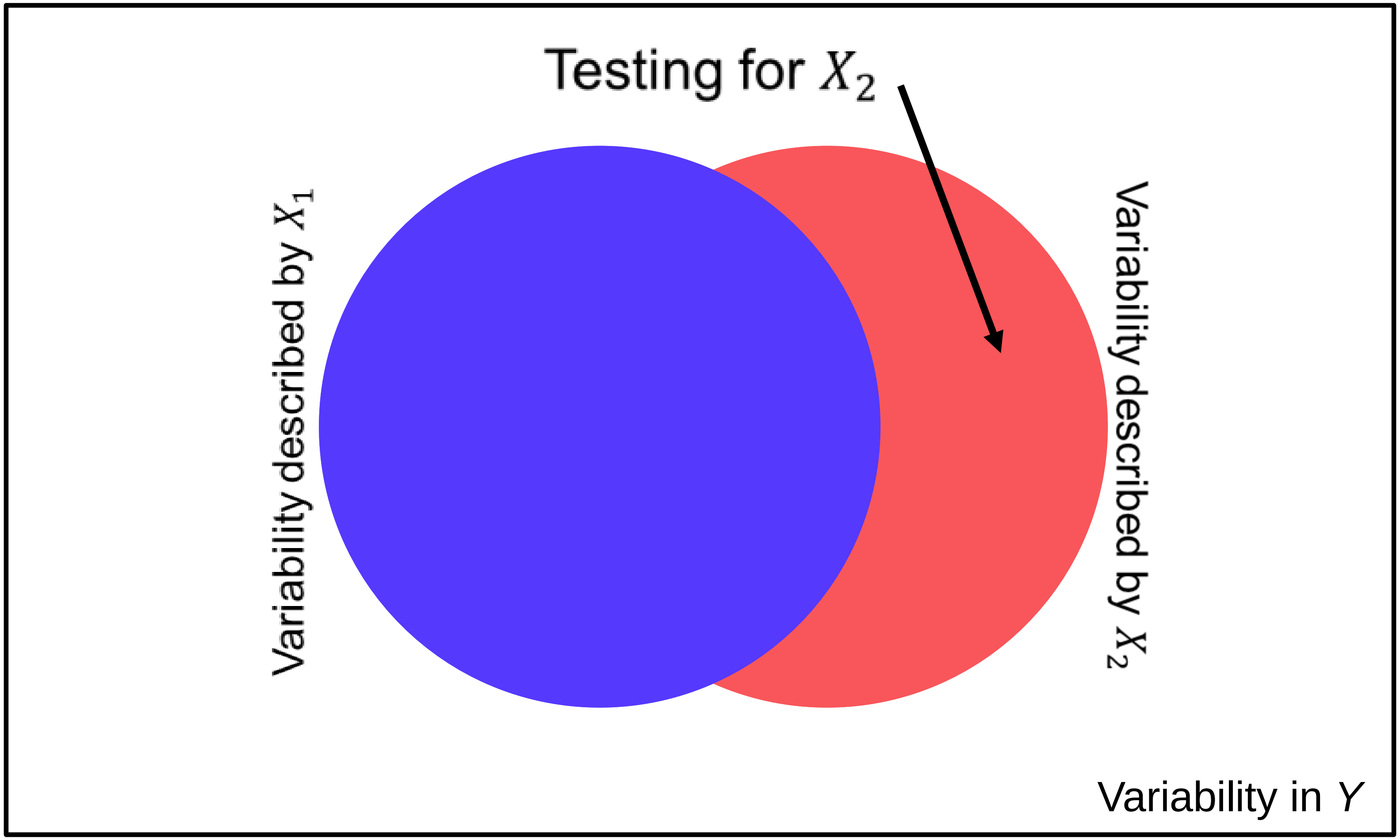
Correlated regressors



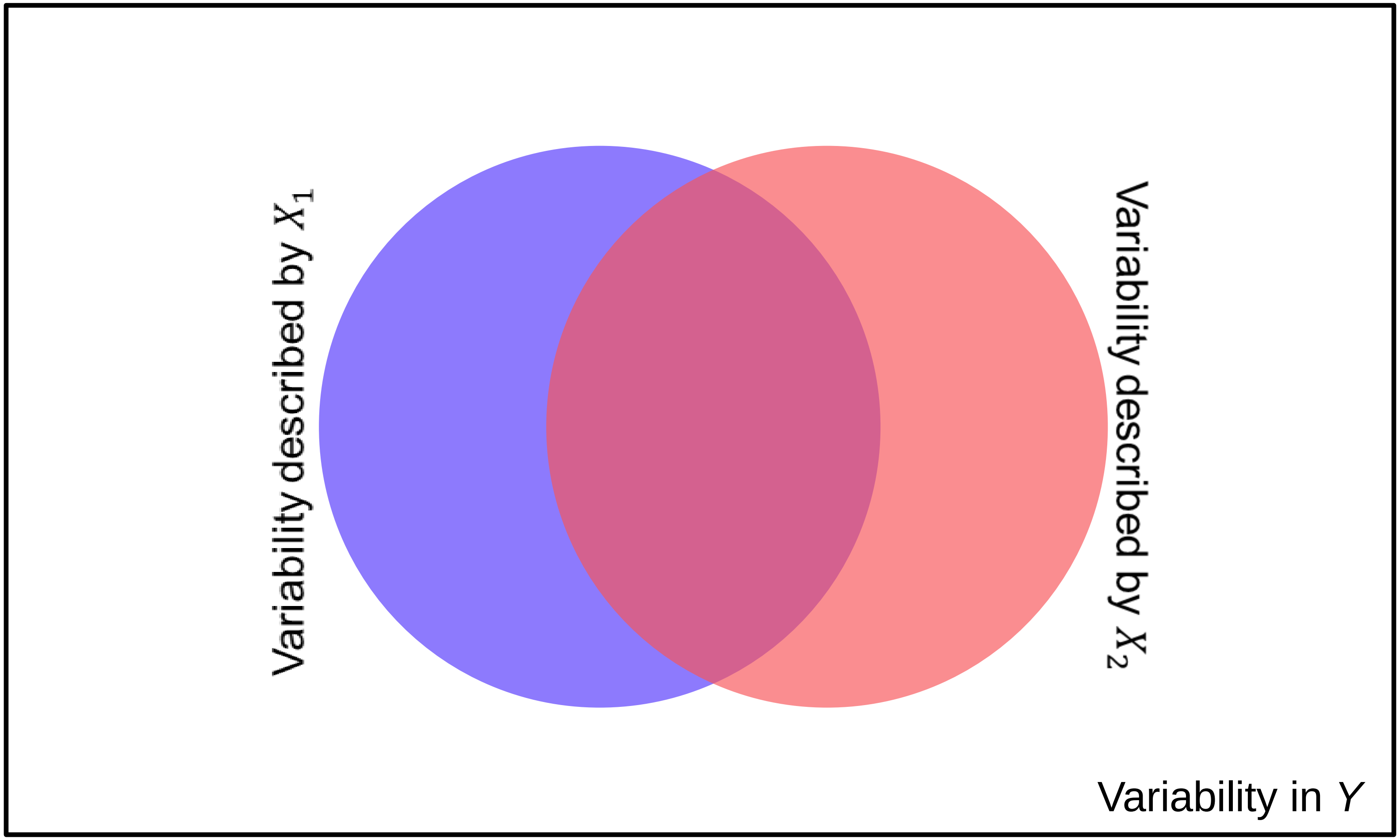
Correlated regressors



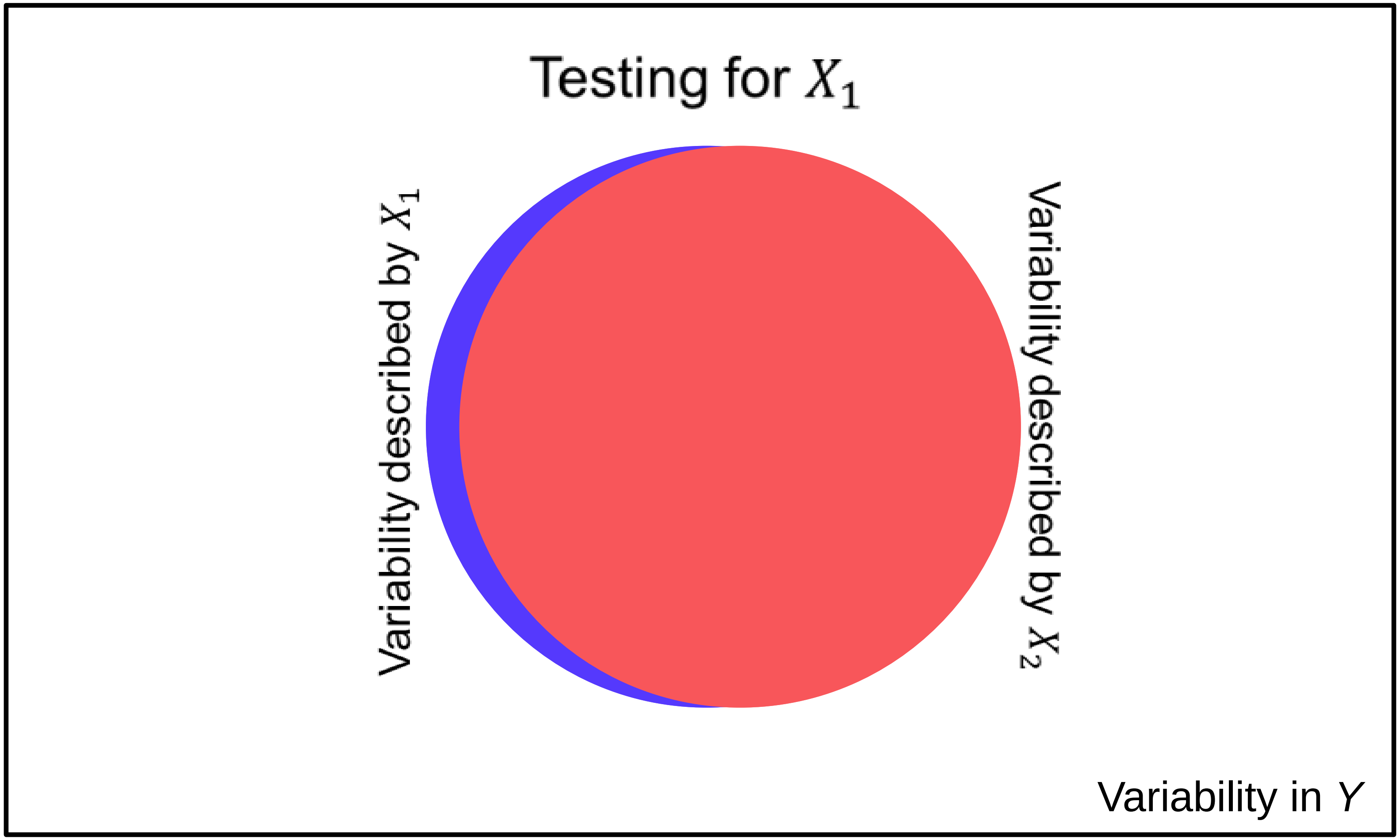
Correlated regressors



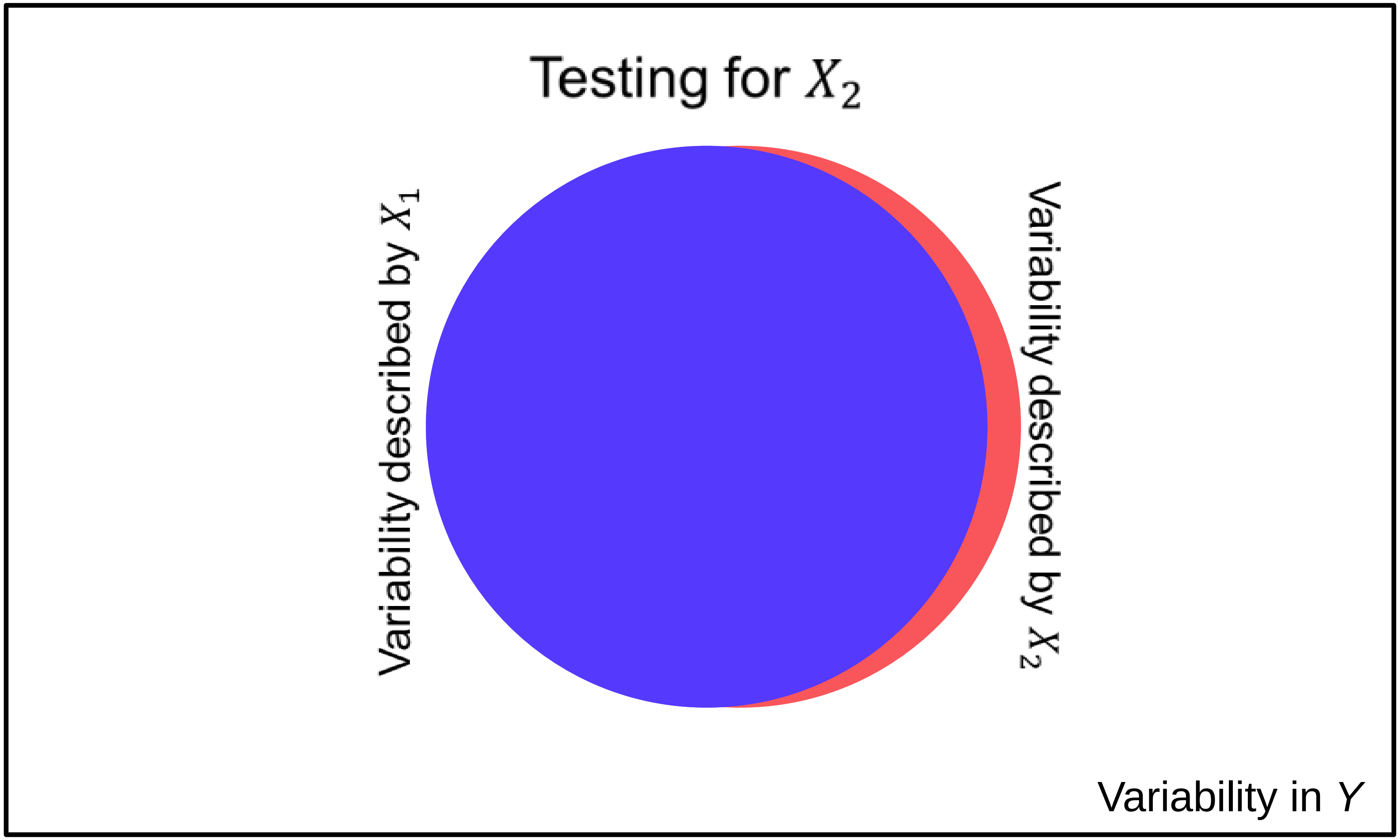
Correlated regressors



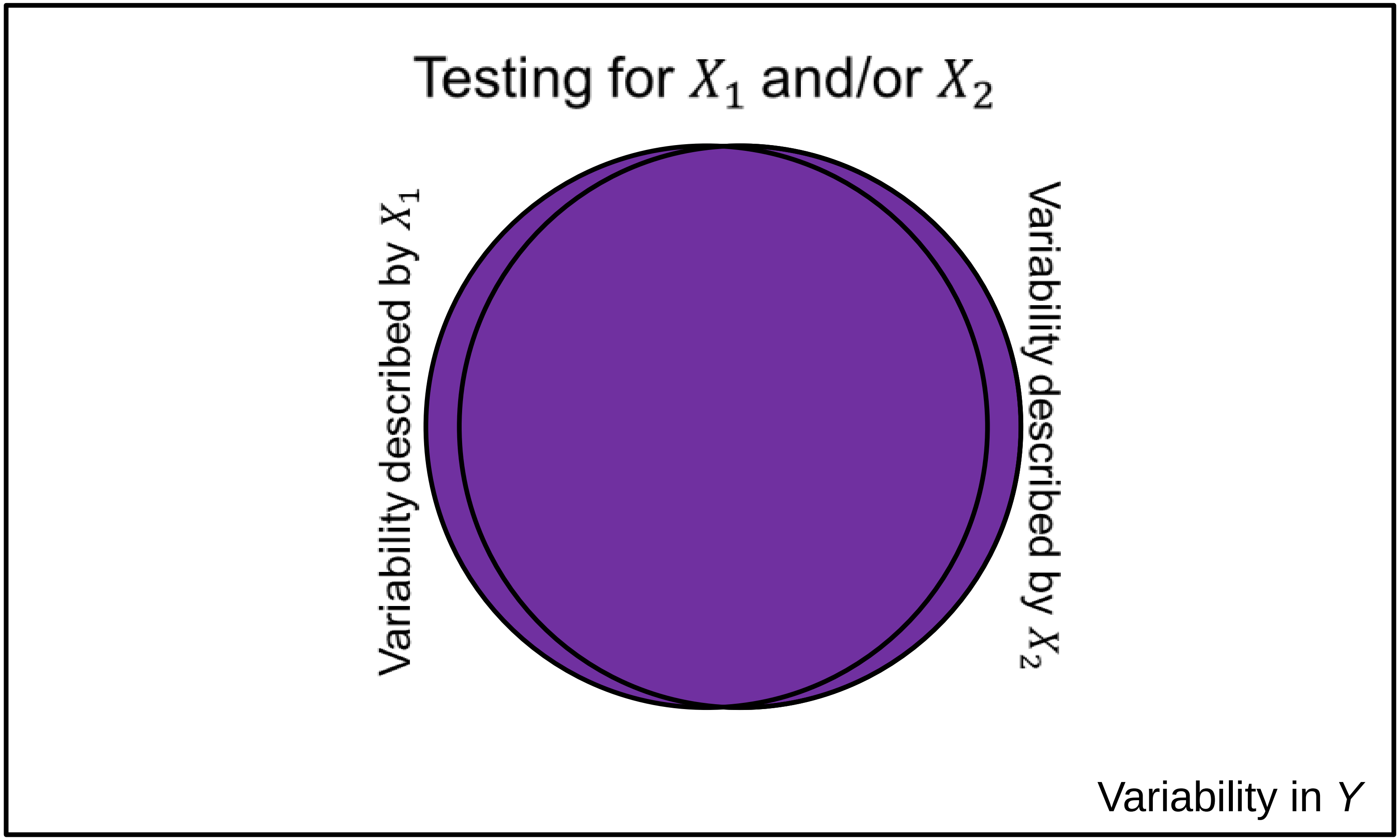
Correlated regressors



Correlated regressors



Correlated regressors



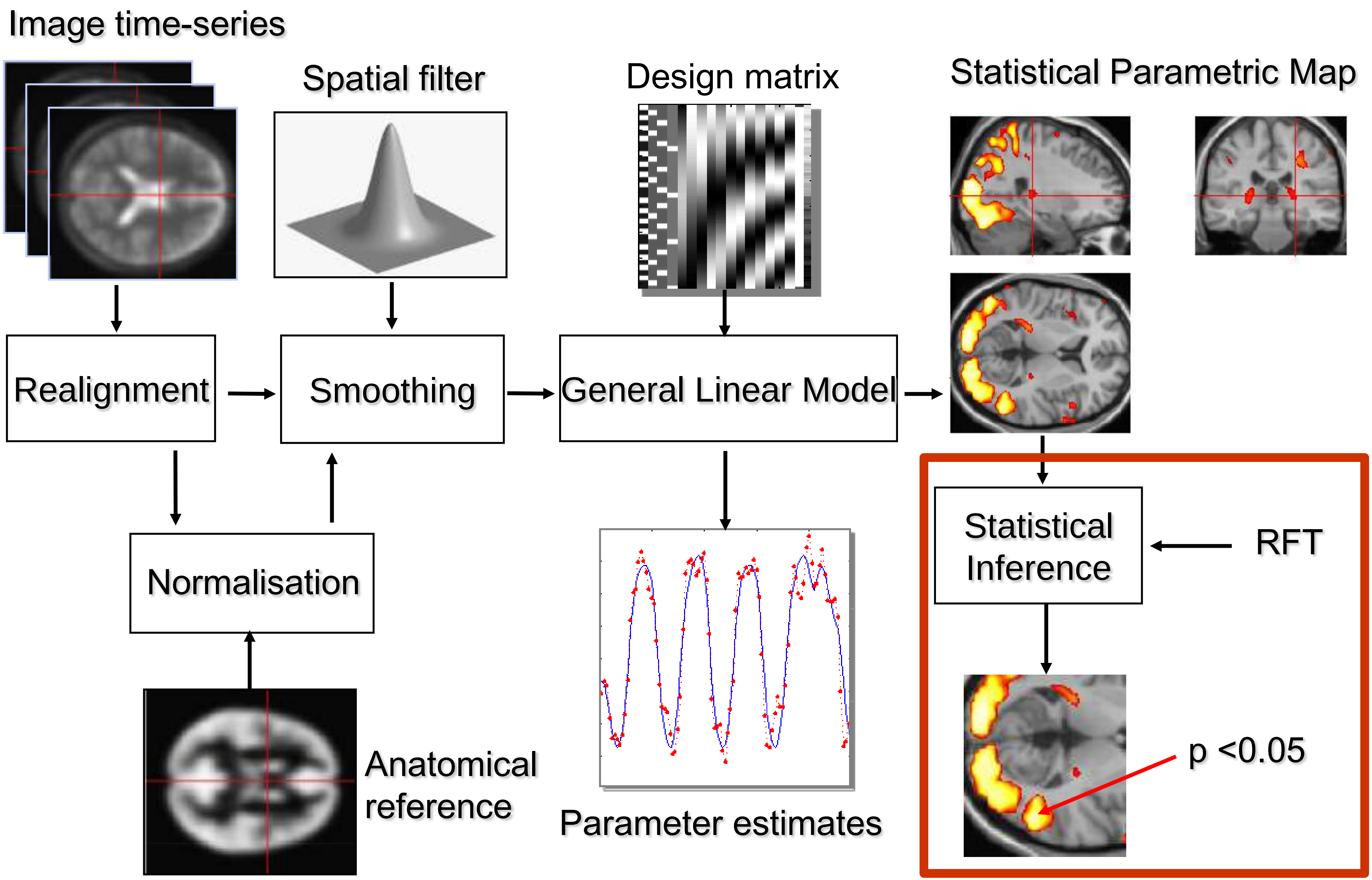
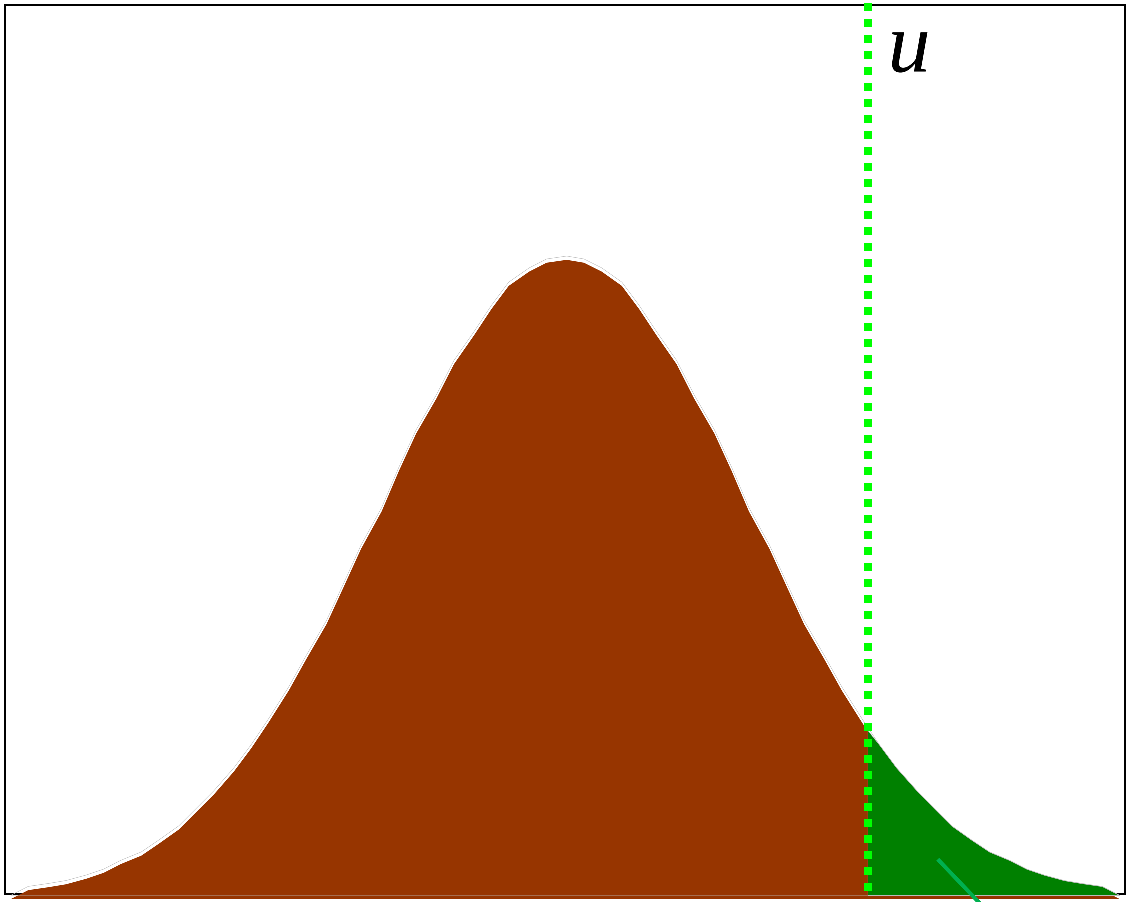


Image time-series
Spatial filter
Design matrix
Statistical Parametric Map
Realignment
Smoothing
General Linear Model
Normalisation
Anatomical reference
Parameter estimates
Statistical Inference
RFT
p <0.05

Inference at a single voxel



Null distribution of test statistic T

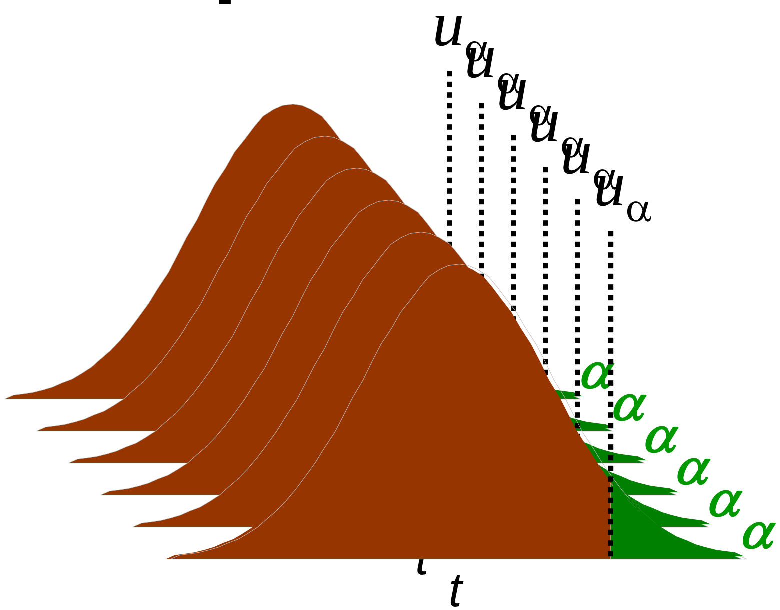
$\hat{\beta}_1 = 3.9831$

Null Hypothesis H_0:
zero activation

Decision rule (threshold) u:
determines false positive rate α

$\Rightarrow$ Choose u to give acceptable α under H_0

Multiple tests

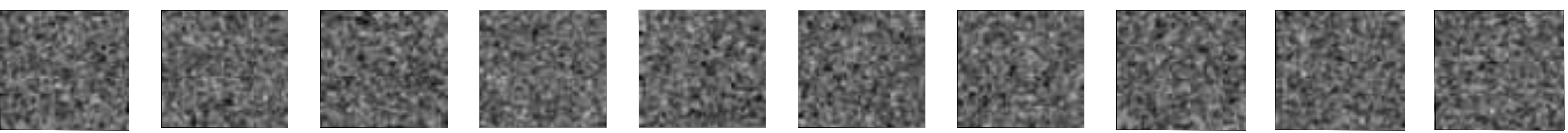


If we have 100,000 voxels,

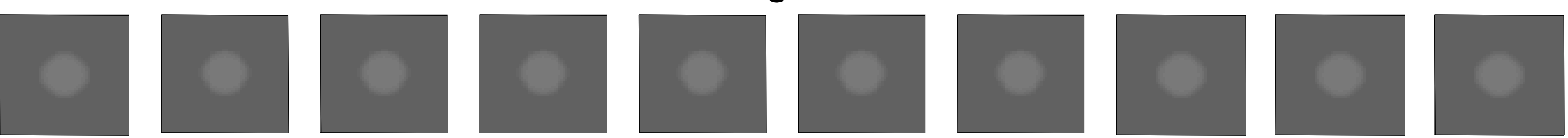
α=0.05 $\Rightarrow$ 5,000 false positive voxels.

This is clearly undesirable; to correct for this we can define a null hypothesis for a collection of tests.

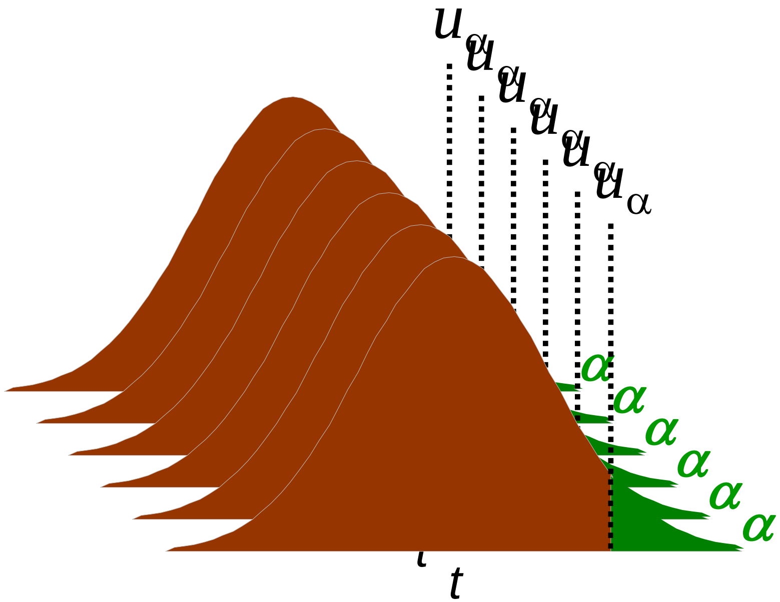
Noise



Signal



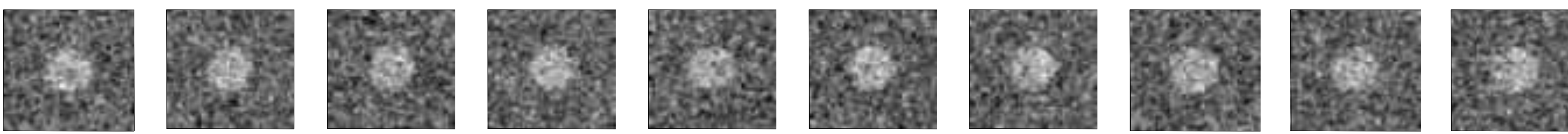
Multiple tests



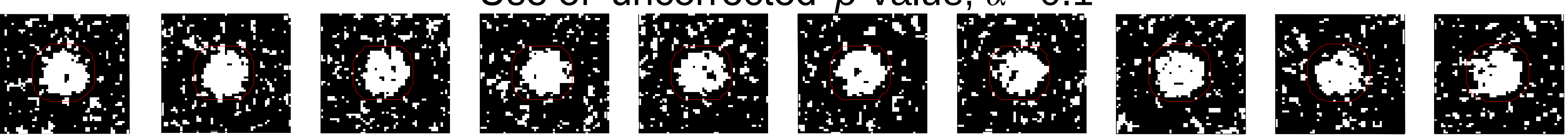
If we have 100,000 voxels,

α=0.05 $\Rightarrow$ 5,000 false positive voxels.

This is clearly undesirable; to correct for this we can define a null hypothesis for a collection of tests.



Use of ‘uncorrected’ *p*-value, α =0.1



11.3% 11.3% 12.5% 10.8% 11.5% 10.0% 10.7% 11.2% 10.2% 9.5%

Percentage of Null Pixels that are False Positives

Family-Wise Null Hypothesis

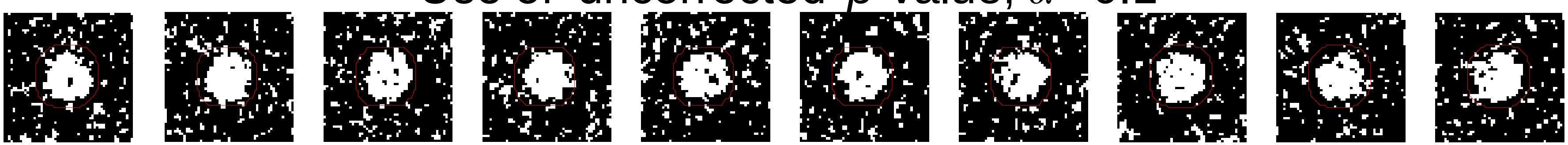
Family-Wise Null Hypothesis:
Activation is zero everywhere

If we reject a voxel null hypothesis at *any* voxel,
we reject the family-wise Null hypothesis

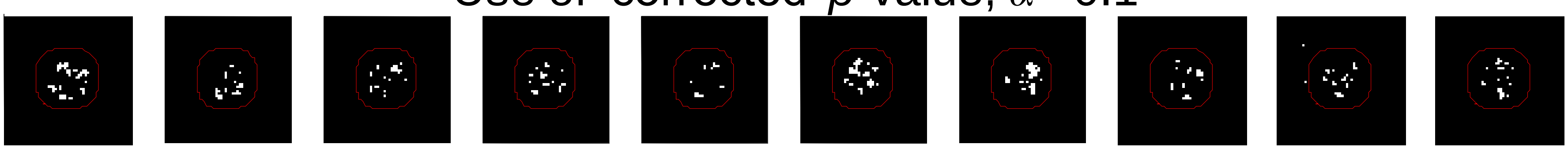
A FP ***anywhere*** in the image gives a **Family Wise Error** (FWE)

Family-Wise Error rate (FWER) = ‘*corrected*’ *p*-value

Use of ‘uncorrected’ *p*-value, α =0.1



Use of ‘corrected’ *p*-value, α =0.1



FWE

Bonferroni correction

The Family-Wise Error rate (FWER), α_{FWE}, for a family of *N* tests follows the inequality:

$$\alpha_{FWE} \leq N\alpha$$

where α is the test-wise error rate.

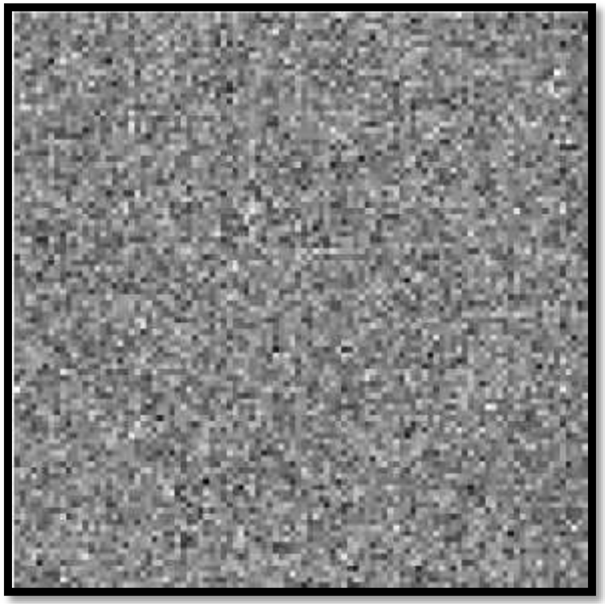
Therefore, to ensure a particular FWER choose:

$$\alpha = \frac{\alpha_{FWE}}{N}$$

This correction does not require the tests to be independent but becomes very stringent if dependence.

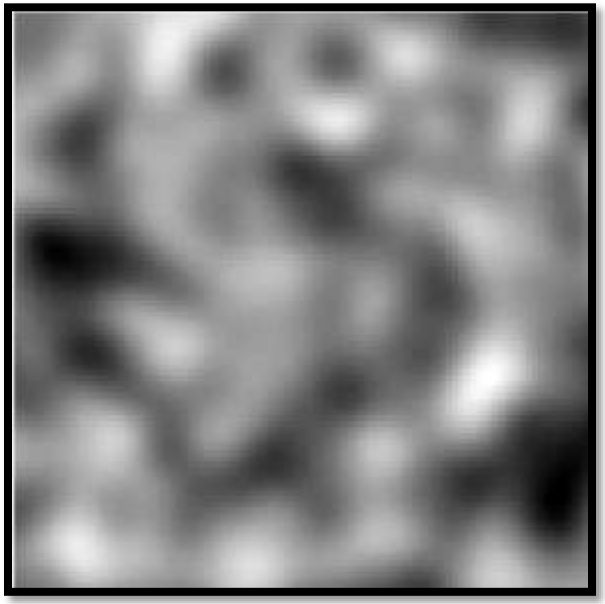
Spatial correlations

100 x 100 independent tests



Discrete data

Spatially correlated tests (FWHM=10)



Spatially extended data

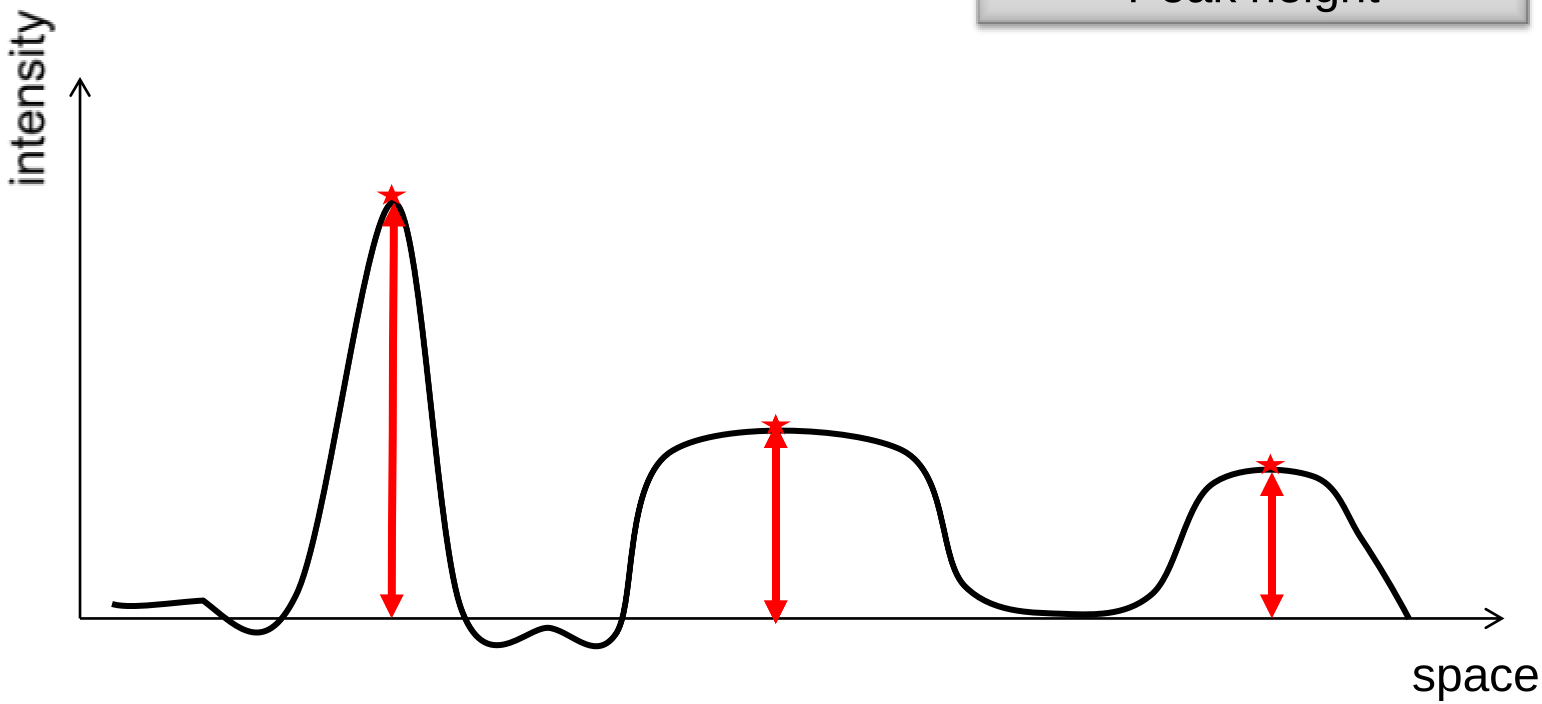
Bonferroni is too conservative for spatial correlated data.

$$10{,}000 \text{ voxels} \Rightarrow \alpha_{BONF} = \frac{0.05}{10{,}000} \Rightarrow u_c = 4.42 \quad (\text{uncorrected } u = 1.64)$$

Topological inference

Peak level inference

Topological feature:
Peak height

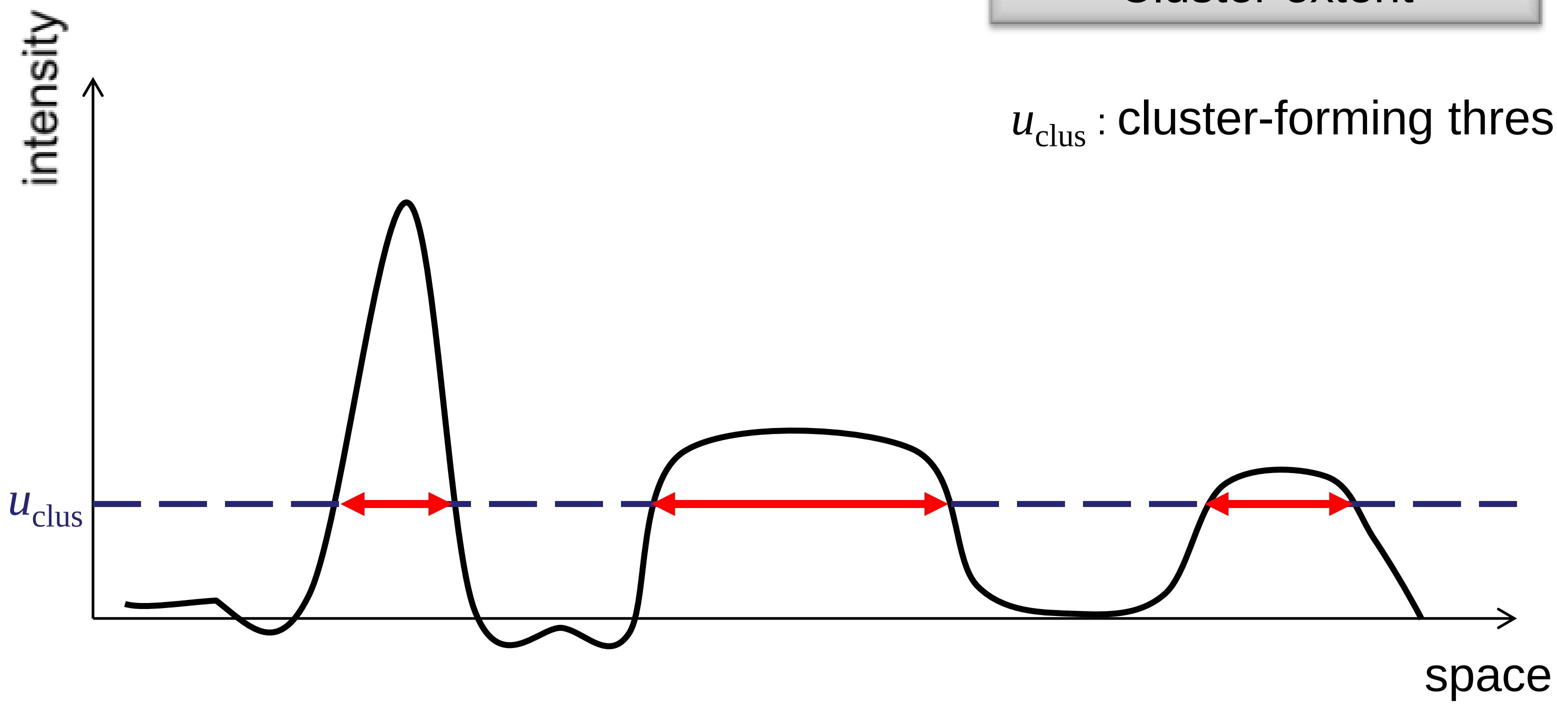


Topological inference

Cluster level inference

Topological feature:
Cluster extent

u_{clus} : cluster-forming threshold

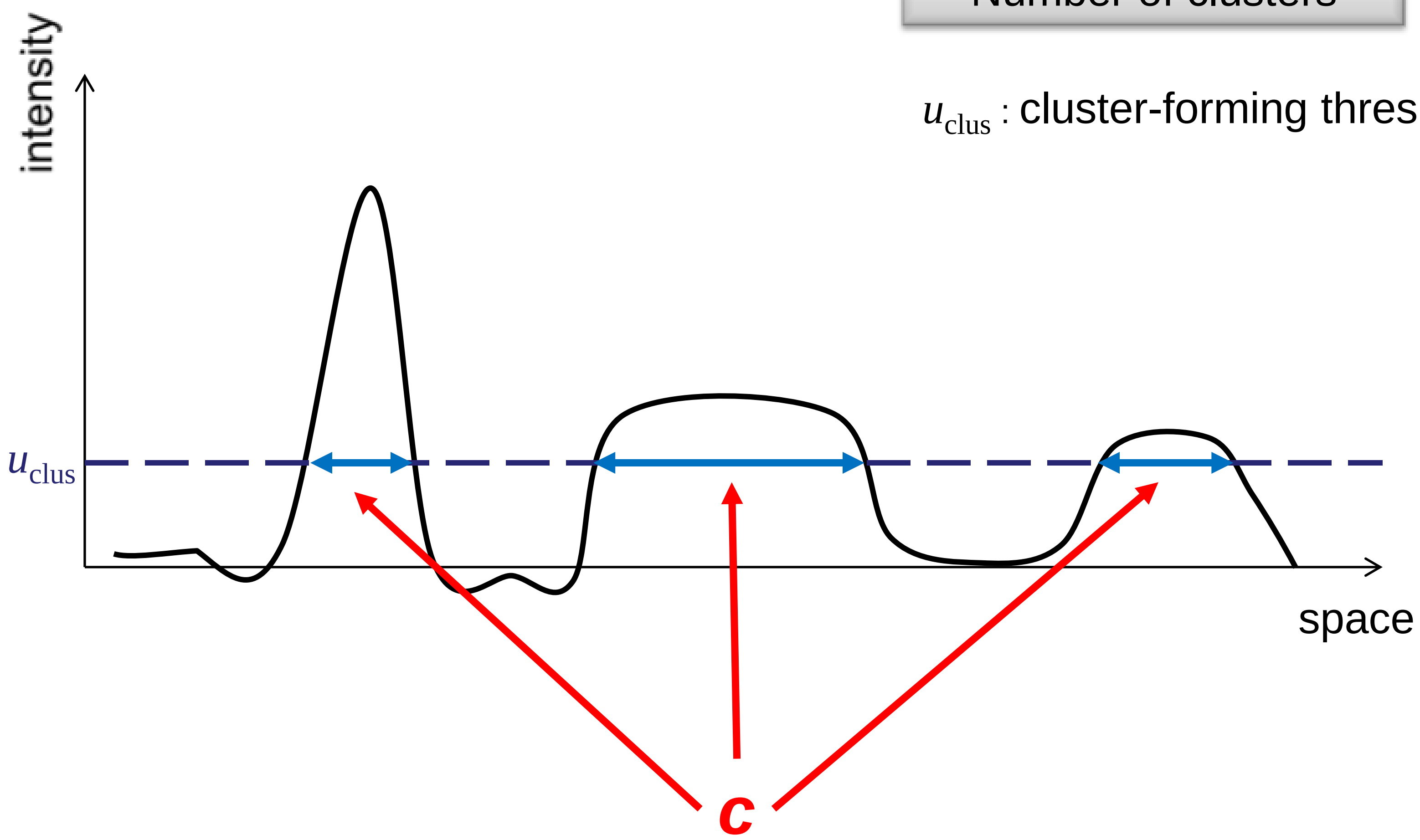


Topological inference

Set level inference

Topological feature:
Number of clusters

u_{clus} : cluster-forming threshold



RFT and Euler Characteristic

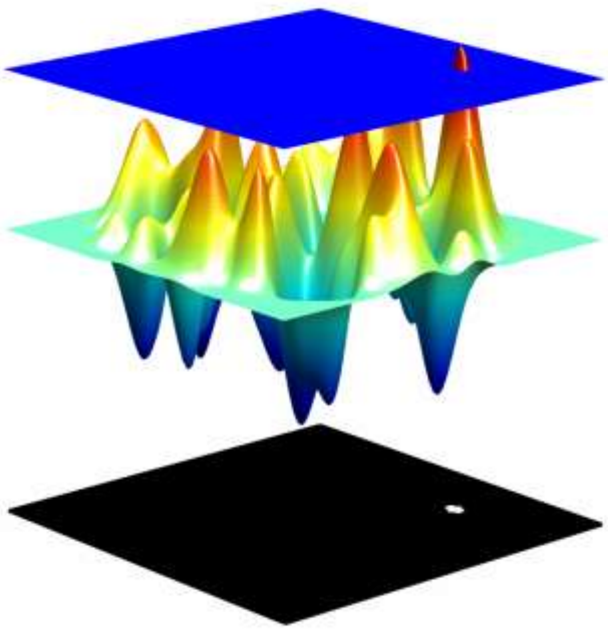
Euler Characteristic χ_u:

- Topological measure

  χ_u = # blobs - # holes

- at high threshold u:

  χ_u = # blobs



$$
\begin{aligned}
FWER &= p(FWE) \\
&\approx E[\chi_u] \\
&\propto \lambda(\Omega)|\Lambda|^{1/2}\, u \exp(-u^2/2)/(2\pi)^{3/2}
\end{aligned}
$$

Search volume

Roughness (1/smoothness)

Threshold

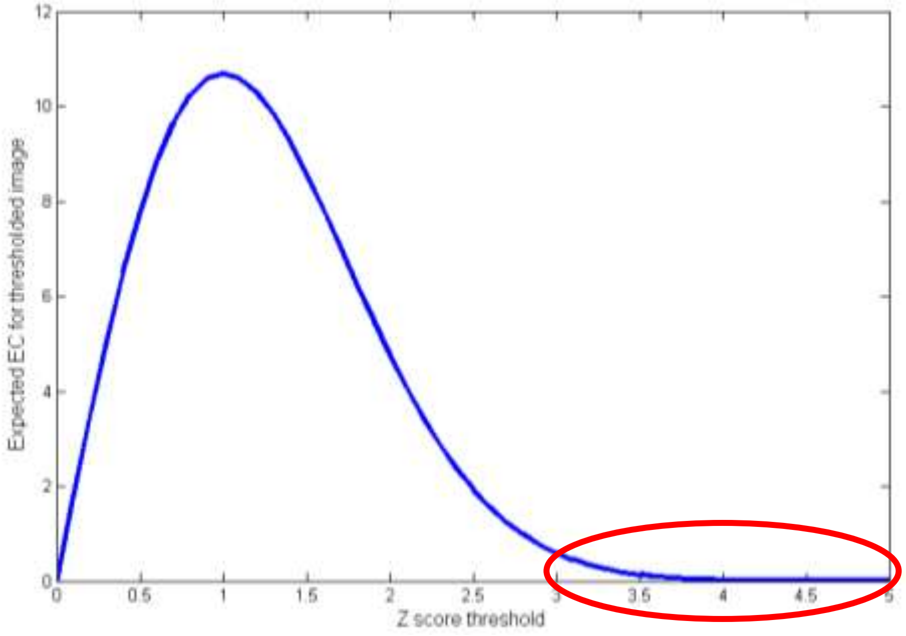
Expected Euler Characteristic

Testing for χ_1

100 x 100 Gaussian Random Field
with FWHM=10 smoothing

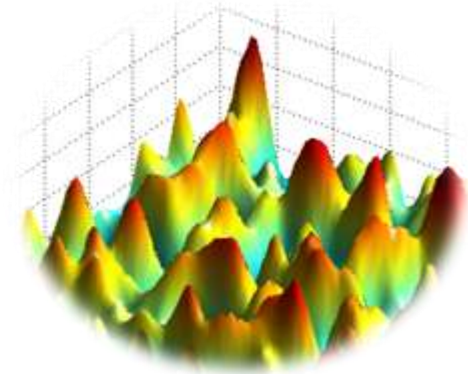
$\alpha_{FWE} = 0.05 \Rightarrow u_{RFT} = 3.8$

$(u_{BONF} = 4.42, u_{uncorr} = 1.64)$

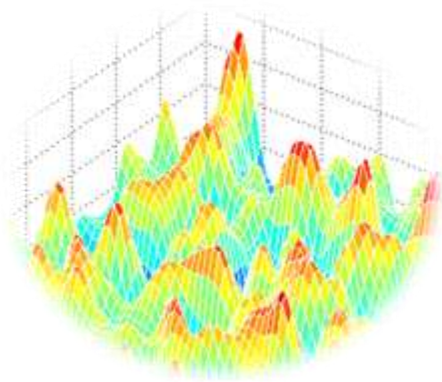


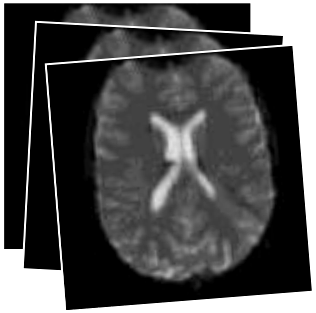
Random Field Theory

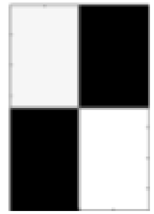
- The statistic image is assumed to be a good lattice representation of an underlying continuous stationary random field.
  Typically, FWHM > 3 voxels
  (combination of intrinsic and extrinsic smoothing)



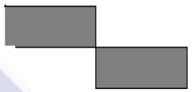
- *A priori* hypothesis about where an activation should be, reduce search volume $\Rightarrow$ Small Volume Correction:
  - mask defined by (probabilistic) anatomical atlases
  - mask defined by separate "functional localisers"
  - mask defined by orthogonal contrasts
  - (spherical) search volume around previously reported coordinates



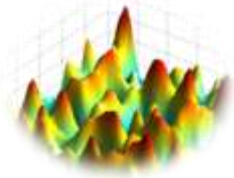


$y =$  $\beta + \varepsilon$

Contrast c



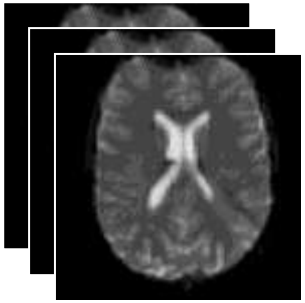
Random Field Theory



Pre-processings

General Linear Model

Statistical Inference



$$\varepsilon \sim N(0, \sigma^2 I)$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{rank(X)}$$